

RUSSIAN FEDERAL COMMITTEE
FOR HIGHER EDUCATION

BASHKIR STATE UNIVERSITY

SHARIPOV R. A.

**COURSE OF LINEAR ALGEBRA
AND MULTIDIMENSIONAL GEOMETRY**

The Textbook

VISIT...

LANZAROTE
Caliente.COM

MSC 97U20
 PACS 01.30.Pp
 UDC 512.64

Sharipov R. A. **Course of Linear Algebra and Multidimensional Geometry**: the textbook / Publ. of Bashkir State University — Ufa, 1996. — pp. 143. ISBN 5-7477-0099-5.

This book is written as a textbook for the course of multidimensional geometry and linear algebra. At Mathematical Department of Bashkir State University this course is taught to the first year students in the Spring semester. It is a part of the basic mathematical education. Therefore, this course is taught at Physical and Mathematical Departments in all Universities of Russia.

In preparing Russian edition of this book I used the computer typesetting on the base of the $\mathcal{A}\mathcal{M}\mathcal{S}$ - $\text{\TeX}$ package and I used the Cyrillic fonts of Lh-family distributed by the CyrTUG association of Cyrillic $\text{\TeX}$ users. English edition of this book is also typeset by means of the $\mathcal{A}\mathcal{M}\mathcal{S}$ - $\text{\TeX}$ package.

Referees: Computational Mathematics and Cybernetics group of Ufa
 State University for Aircraft and Technology (УГАТУ);
 Prof. S. I. Pinchuk, Chelyabinsk State University for Technol-
 ogy (ЧГТУ) and Indiana University.

Contacts to author.

Office: Mathematics Department, Bashkir State University,
 32 Frunze street, 450074 Ufa, Russia
 Phone: 7-(3472)-23-67-18
 Fax: 7-(3472)-23-67-74
 Home: 5 Rabochaya street, 450003 Ufa, Russia
 Phone: 7-(917)-75-55-786
 E-mails: R.Sharipov@ic.bashedu.ru
r-sharipov@mail.ru
ra_sharipov@lycos.com
ra_sharipov@hotmail.com
 URL: <http://www.geocities.com/r-sharipov>

CONTENTS.

CONTENTS.	3.
PREFACE.	5.
CHAPTER I. LINEAR VECTOR SPACES AND LINEAR MAPPINGS.	6.
§ 1. The sets and mappings.	6.
§ 2. Linear vector spaces.	10.
§ 3. Linear dependence and linear independence.	14.
§ 4. Spanning systems and bases.	18.
§ 5. Coordinates. Transformation of the coordinates of a vector under a change of basis.	22.
§ 6. Intersections and sums of subspaces.	27.
§ 7. Cosets of a subspace. The concept of factorspace.	31.
§ 8. Linear mappings.	36.
§ 9. The matrix of a linear mapping.	39.
§ 10. Algebraic operations with mappings. The space of homomorphisms $\text{Hom}(V, W)$	45.
CHAPTER II. LINEAR OPERATORS.	50.
§ 1. Linear operators. The algebra of endomorphisms $\text{End}(V)$ and the group of automorphisms $\text{Aut}(V)$	50.
§ 2. Projection operators.	56.
§ 3. Invariant subspaces. Restriction and factorization of operators.	61.
§ 4. Eigenvalues and eigenvectors.	66.
§ 5. Nilpotent operators.	72.
§ 6. Root subspaces. Two theorems on the sum of root subspaces.	79.
§ 7. Jordan basis of a linear operator. Hamilton-Cayley theorem.	83.
CHAPTER III. DUAL SPACE.	87.
§ 1. Linear functionals. Vectors and covectors. Dual space.	87.
§ 2. Transformation of the coordinates of a covector under a change of basis.	92.
§ 3. Orthogonal complements in a dual spaces.	94.
§ 4. Conjugate mapping.	97.
CHAPTER IV. BILINEAR AND QUADRATIC FORMS.	100.
§ 1. Symmetric bilinear forms and quadratic forms. Recovery formula.	100.
§ 2. Orthogonal complements with respect to a quadratic form.	103.

§ 3. Transformation of a quadratic form to its canonic form. Inertia indices and signature.	108.
§ 4. Positive quadratic forms. Sylvester's criterion.	114.
CHAPTER V. EUCLIDEAN SPACES.	119.
§ 1. The norm and the scalar product. The angle between vectors. Orthonormal bases.	119.
§ 2. Quadratic forms in a Euclidean space. Diagonalization of a pair of quadratic forms.	123.
§ 3. Selfadjoint operators. Theorem on the spectrum and the basis of eigenvectors for a selfadjoint operator.	127.
§ 4. Isometries and orthogonal operators.	132.
CHAPTER VI. AFFINE SPACES.	136.
§ 1. Points and parallel translations. Affine spaces.	136.
§ 2. Euclidean point spaces. Quadrics in a Euclidean space.	139.
REFERENCES.	143.

PREFACE.

There are two approaches to stating the linear algebra and the multidimensional geometry. The first approach can be characterized as the «coordinates and matrices approach». The second one is the «invariant geometric approach».

In most of textbooks the coordinates and matrices approach is used. It starts with considering the systems of linear algebraic equations. Then the theory of determinants is developed, the matrix algebra and the geometry of the space $\mathbb{R}^n$ are considered. This approach is convenient for initial introduction to the subject since it is based on very simple concepts: the numbers, the sets of numbers, the numeric matrices, linear functions, and linear equations. The proofs within this approach are conceptually simple and mostly are based on calculations. However, in further statement of the subject the coordinates and matrices approach is not so advantageous. Computational proofs become huge, while the intension to consider only numeric objects prevents us from introducing and using new concepts.

The invariant geometric approach, which is used in this book, starts with the definition of abstract linear vector space. Thereby the coordinate representation of vectors is not of crucial importance; the set-theoretic methods commonly used in modern algebra become more important. Linear vector space is the very object to which these methods apply in a most simple and effective way: proofs of many facts can be shortened and made more elegant.

The invariant geometric approach lets the reader to get prepared to the study of more advanced branches of mathematics such as differential geometry, commutative algebra, algebraic geometry, and algebraic topology. I prefer a self-sufficient way of explanation. The reader is assumed to have only minimal preliminary knowledge in matrix algebra and in theory of determinants. This material is usually given in courses of general algebra and analytic geometry.

Under the term «numeric field» in this book we assume one of the following three fields: the field of rational numbers $\mathbb{Q}$, the field of real numbers $\mathbb{R}$, or the field of complex numbers $\mathbb{C}$. Therefore the reader should not know the general theory of numeric fields.

I am grateful to E. B. Rudenko for reading and correcting the manuscript of Russian edition of this book.

May, 1996;
May, 2004.

R. A. Sharipov.

CHAPTER I

LINEAR VECTOR SPACES AND LINEAR MAPPINGS.

§ 1. The sets and mappings.

The concept of a *set* is a basic concept of modern mathematics. It denotes any group of objects for some reasons distinguished from other objects and grouped together. Objects constituting a given set are called *the elements* of this set. We usually assign some literal names (identifiers) to the sets and to their elements. Suppose the set A consists of three objects m , n , and q . Then we write

$$A = \{m, n, q\}.$$

The fact that m is an element of the set A is denoted by the membership sign: $m \in A$. The writing $p \notin A$ means that the object p is not an element of the set A .

If we have several sets, we can gather all of their elements into one set which is called the *union* of initial sets. In order to denote this gathering operation we use the union sign $\cup$. If we gather the elements each of which belongs to all of our sets, they constitute a new set which is called the *intersection* of initial sets. In order to denote this operation we use the intersection sign $\cap$.

If a set A is a part of another set B , we denote this fact as $A \subset B$ or $A \subseteq B$ and say that the set A is a *subset* of the set B . Two signs $\subset$ and $\subseteq$ are equivalent. However, using the sign $\subseteq$, we emphasize that the condition $A \subset B$ does not exclude the coincidence of sets $A = B$. If $A \subsetneq B$, then we say that the set A is a *strict subset* in the set B .

The term *empty set* is used to denote the set $\emptyset$ that comprises no elements at all. The empty set is assumed to be a part of any set: $\emptyset \subset A$.

DEFINITION 1.1. The mapping $f: X \rightarrow Y$ from the set X to the set Y is a rule f applicable to any element x of the set X and such that, being applied to a particular element $x \in X$, uniquely defines some element $y = f(x)$ in the set Y .

The set X in the definition 1.1 is called *the domain* of the mapping f . The set Y in the definition 1.1 is called *the domain of values* of the mapping f . The writing $f(x)$ means that the rule f is applied to the element x of the set X . The element $y = f(x)$ obtained as a result of applying f to x is called *the image* of x under the mapping f .

Let A be a subset of the set X . The set $f(A)$ composed by the images of all elements $x \in A$ is called *the image* of the subset A under the mapping f :

$$f(A) = \{y \in Y: \exists x ((x \in A) \ \& \ (f(x) = y))\}.$$

If $A = X$, then the image $f(X)$ is called *the image of the mapping* f . There is special notation for this image: $f(X) = \text{Im } f$. *The set of values* is another term used for denoting $\text{Im } f = f(X)$; don't confuse it with *the domain of values*.

Let y be an element of the set Y . Let's consider the set $f^{-1}(y)$ consisting of all elements $x \in X$ that are mapped to the element y . This set $f^{-1}(y)$ is called *the total preimage* of the element y :

$$f^{-1}(y) = \{x \in X : f(x) = y\}.$$

Suppose that B is a subset in Y . Taking the union of total preimages for all elements of the set B , we get *the total preimage* of the set B itself:

$$f^{-1}(B) = \{x \in X : f(x) \in B\}.$$

It is clear that for the case $B = Y$ the total preimage $f^{-1}(Y)$ coincides with X . Therefore there is no special sign for denoting $f^{-1}(Y)$.

DEFINITION 1.2. The mapping $f: X \rightarrow Y$ is called *injective* if images of any two distinct elements $x_1 \neq x_2$ are different, i.e. $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$.

DEFINITION 1.3. The mapping $f: X \rightarrow Y$ is called *surjective* if total preimage $f^{-1}(y)$ of any element $y \in Y$ is not empty.

DEFINITION 1.4. The mapping $f: X \rightarrow Y$ is called a *bijective* mapping or a *one-to-one* mapping if total preimage $f^{-1}(y)$ of any element $y \in Y$ is a set consisting of exactly one element.

THEOREM 1.1. *The mapping $f: X \rightarrow Y$ is bijective if and only if it is injective and surjective simultaneously.*

PROOF. According to the statement of theorem 1.1, simultaneous injectivity and surjectivity is necessary and sufficient condition for bijectivity of the mapping $f: X \rightarrow Y$. Let's prove the necessity of this condition for the beginning.

Suppose that the mapping $f: X \rightarrow Y$ is bijective. Then for any $y \in Y$ the total preimage $f^{-1}(y)$ consists of exactly one element. This means that it is not empty. This fact proves the surjectivity of the mapping $f: X \rightarrow Y$.

However, we need to prove that f is not only surjective, but bijective as well. Let's prove the bijectivity of f by contradiction. If the mapping f is not bijective, then there are two distinct elements $x_1 \neq x_2$ in X such that $f(x_1) = f(x_2)$. Let's denote $y = f(x_1) = f(x_2)$ and consider the total preimage $f^{-1}(y)$. From the equality $f(x_1) = y$ we derive $x_1 \in f^{-1}(y)$. Similarly from $f(x_2) = y$ we derive $x_2 \in f^{-1}(y)$. Hence, the total preimage $f^{-1}(y)$ is a set containing at least two distinct elements x_1 and x_2 . This fact contradicts the bijectivity of the mapping $f: X \rightarrow Y$. Due to this contradiction we conclude that f is surjective and injective simultaneously. Thus, we have proved the necessity of the condition stated in theorem 1.1.

Let's proceed to the proof of sufficiency. Suppose that the mapping $f: X \rightarrow Y$ is injective and surjective simultaneously. Due to the surjectivity the sets $f^{-1}(y)$ are non-empty for all $y \in Y$. Suppose that someone of them contains more than one element. If $x_1 \neq x_2$ are two distinct elements of the set $f^{-1}(y)$, then $f(x_1) = y = f(x_2)$. However, this equality contradicts the injectivity of the mapping $f: X \rightarrow Y$. Hence, each set $f^{-1}(y)$ is non-empty and contains exactly one element. Thus, we have proved the bijectivity of the mapping f . $\square$

THEOREM 1.2. *The mapping $f: X \rightarrow Y$ is surjective if and only if $\text{Im } f = Y$.*

PROOF. If the mapping $f: X \rightarrow Y$ is surjective, then for any element $y \in Y$ the total preimage $f^{-1}(y)$ is not empty. Choosing some element $x \in f^{-1}(y)$, we get $y = f(x)$. Hence, each element $y \in Y$ is an image of some element x under the mapping f . This proves the equality $\text{Im } f = Y$.

Conversely, if $\text{Im } f = Y$, then any element $y \in Y$ is an image of some element $x \in X$, i.e. $y = f(x)$. Hence, for any $y \in Y$ the total preimage $f^{-1}(y)$ is not empty. This means that f is a surjective mapping. $\square$

Let's consider two mappings $f: X \rightarrow Y$ and $g: Y \rightarrow Z$. Choosing an arbitrary element $x \in X$ we can apply f to it. As a result we get the element $f(x) \in Y$. Then we can apply g to $f(x)$. The successive application of two mappings $g(f(x))$ yields a rule that associates each element $x \in X$ with some uniquely determined element $z = g(f(x)) \in Z$, i.e. we have a mapping $\varphi: X \rightarrow Z$. This mapping is called *the composition* of two mappings f and g . It is denoted as $\varphi = g \circ f$.

THEOREM 1.3. *The composition $g \circ f$ of two injective mappings $f: X \rightarrow Y$ and $g: Y \rightarrow Z$ is an injective mapping.*

PROOF. Let's consider two elements x_1 and x_2 of the set X . Denote $y_1 = f(x_1)$ and $y_2 = f(x_2)$. Therefore $g \circ f(x_1) = g(y_1)$ and $g \circ f(x_2) = g(y_2)$. Due to the injectivity of f from $x_1 \neq x_2$ we derive $y_1 \neq y_2$. Then due to the injectivity of g from $y_1 \neq y_2$ we derive $g(y_1) \neq g(y_2)$. Hence, $g \circ f(x_1) \neq g \circ f(x_2)$. The injectivity of the composition $g \circ f$ is proved. $\square$

THEOREM 1.4. *The composition $g \circ f$ of two surjective mappings $f: X \rightarrow Y$ and $g: Y \rightarrow Z$ is a surjective mapping.*

PROOF. Let's take an arbitrary element $z \in Z$. Due to the surjectivity of g the total preimage $g^{-1}(z)$ is not empty. Let's choose some arbitrary vector $y \in g^{-1}(z)$ and consider its total preimage $f^{-1}(y)$. Due to the surjectivity of f it is not empty. Then choosing an arbitrary vector $x \in f^{-1}(y)$, we get $g \circ f(x) = g(f(x)) = g(y) = z$. This means that $x \in (g \circ f)^{-1}(z)$. Hence, the total preimage $(g \circ f)^{-1}(z)$ is not empty. The surjectivity of $g \circ f$ is proved. $\square$

As an immediate consequence of the above two theorems we obtain the following theorem on composition of two bijections.

THEOREM 1.5. *The composition $g \circ f$ of two bijective mappings $f: X \rightarrow Y$ and $g: Y \rightarrow Z$ is a bijective mapping.*

Let's consider three mappings $f: X \rightarrow Y$, $g: Y \rightarrow Z$, and $h: Z \rightarrow U$. Then we can form two different compositions of these mappings:

$$\varphi = h \circ (g \circ f), \quad \psi = (h \circ g) \circ f. \quad (1.1)$$

The fact of coincidence of these two mappings is formulated as the following theorem on associativity.

THEOREM 1.6. *The operation of composition for the mappings is an associative operation, i.e. $h \circ (g \circ f) = (h \circ g) \circ f$.*

PROOF. According to the definition 1.1, the coincidence of two mappings $\varphi: X \rightarrow U$ and $\psi: X \rightarrow U$ is verified by verifying the equality $\varphi(x) = \psi(x)$ for an arbitrary element $x \in X$. Let's denote $\alpha = h \circ g$ and $\beta = g \circ f$. Then

$$\begin{aligned}\varphi(x) &= h \circ \beta(x) = h(\beta(x)) = h(g(f(x))), \\ \psi(x) &= \alpha \circ f(x) = \alpha(f(x)) = h(g(f(x))).\end{aligned}\tag{1.2}$$

Comparing right hand sides of the equalities (1.2), we derive the required equality $\varphi(x) = \psi(x)$ for the mappings (1.1). Hence, $h \circ (g \circ f) = (h \circ g) \circ f$. $\square$

Let's consider a mapping $f: X \rightarrow Y$ and the pair of identical mappings $\text{id}_X: X \rightarrow X$ and $\text{id}_Y: Y \rightarrow Y$. The last two mappings are defined as follows:

$$\text{id}_X(x) = x, \quad \text{id}_Y(y) = y.$$

DEFINITION 1.5. A mapping $l: Y \rightarrow X$ is called *left inverse* to the mapping $f: X \rightarrow Y$ if $l \circ f = \text{id}_X$.

DEFINITION 1.6. A mapping $r: Y \rightarrow X$ is called *right inverse* to the mapping $f: X \rightarrow Y$ if $f \circ r = \text{id}_Y$.

The problem of existence of the left and right inverse mappings is solved by the following two theorems.

THEOREM 1.7. A mapping $f: X \rightarrow Y$ possesses the left inverse mapping l if and only if it is injective.

THEOREM 1.8. A mapping $f: X \rightarrow Y$ possesses the right inverse mapping r if and only if it is surjective.

PROOF OF THE THEOREM 1.7. Suppose that the mapping f possesses the left inverse mapping l . Let's choose two vectors x_1 and x_2 in the space X and let's denote $y_1 = f(x_1)$ and $y_2 = f(x_2)$. The equality $l \circ f = \text{id}_X$ yields $x_1 = l(y_1)$ and $x_2 = l(y_2)$. Hence, the equality $y_1 = y_2$ implies $x_1 = x_2$ and $x_1 \neq x_2$ implies $y_1 \neq y_2$. Thus, assuming the existence of left inverse mapping l , we deduce that the direct mapping f is injective.

Conversely, suppose that f is an injective mapping. First of all let's choose and fix some element $x_0 \in X$. Then let's consider an arbitrary element $y \in \text{Im } f$. Its total preimage $f^{-1}(y)$ is not empty. For any $y \in \text{Im } f$ we can choose and fix some element $x_y \in f^{-1}(y)$ in non-empty set $f^{-1}(y)$. Then we define the mapping $l: Y \rightarrow X$ by the following equality:

$$l(y) = \begin{cases} x_y & \text{for } y \in \text{Im } f, \\ x_0 & \text{for } y \notin \text{Im } f. \end{cases}$$

Let's study the composition $l \circ f$. It is easy to see that for any $x \in X$ and for $y = f(x)$ the equality $l \circ f(x) = x_y$ is fulfilled. Then $f(x_y) = y = f(x)$. Taking into account the injectivity of f , we get $x_y = x$. Hence, $l \circ f(x) = x$ for any $x \in X$. The equality $l \circ f = \text{id}_X$ for the mapping l is proved. Therefore, this mapping is a required left inverse mapping for f . Theorem is proved. $\square$

PROOF OF THE THEOREM 1.8. Suppose that the mapping f possesses the right inverse mapping r . For an arbitrary element $y \in Y$, from the equality $f \circ r = \text{id}_Y$

we derive $y = f(r(y))$. This means that $r(y) \in f^{-1}(y)$, therefore, the total preimage $f^{-1}(y)$ is not empty. Thus, the surjectivity of f is proved.

Now, conversely, let's assume that f is surjective. Then for any $y \in Y$ the total preimage $f^{-1}(y)$ is not empty. In each non-empty set $f^{-1}(y)$ we choose and mark exactly one element $x_y \in f^{-1}(y)$. Then we can define a mapping by setting $r(y) = x_y$. Since $f(x_y) = y$, we get $f(r(y)) = y$ and $f \circ r = \text{id}_Y$. The existence of the right inverse mapping r for f is established. $\square$

Note that the mappings $l: Y \rightarrow X$ and $r: Y \rightarrow X$ constructed when proving theorems 1.7 and 1.8 in general are not unique. Even the method of constructing them contains definite extent of arbitrariness.

DEFINITION 1.7. A mapping $f^{-1}: Y \rightarrow X$ is called *bilateral inverse mapping* or simply *inverse mapping* for the mapping $f: X \rightarrow Y$ if

$$f^{-1} \circ f = \text{id}_X, \quad f \circ f^{-1} = \text{id}_Y. \quad (1.3)$$

THEOREM 1.9. A mapping $f: X \rightarrow Y$ possesses both left and right inverse mappings l and r if and only if it is bijective. In this case the mappings l and r are uniquely determined. They coincide with each other thus determining the unique bilateral inverse mapping $l = r = f^{-1}$.

PROOF. The first proposition of the theorem 1.9 follows from theorems 1.7, 1.8, and 1.1. Let's prove the remaining propositions of this theorem 1.9. The coincidence $l = r$ is derived from the following chain of equalities:

$$l = l \circ \text{id}_Y = l \circ (f \circ r) = (l \circ f) \circ r = \text{id}_X \circ r = r.$$

The uniqueness of left inverse mapping also follows from the same chain of equalities. Indeed, if we assume that there is another left inverse mapping l' , then from $l = r$ and $l' = r$ it follows that $l = l'$.

In a similar way, assuming the existence of another right inverse mapping r' , we get $l = r$ and $l = r'$. Hence, $r = r'$. Coinciding with each other, the left and right inverse mappings determine the unique bilateral inverse mapping $f^{-1} = l = r$ satisfying the equalities (1.3). $\square$

§ 2. Linear vector spaces.

Let M be a set. *Binary algebraic operation in M* is a rule that maps each ordered pair of elements x, y of the set M to some uniquely determined element $z \in M$. This rule can be denoted as a function $z = f(x, y)$. This notation is called a *prefix* notation for an algebraic operation: the operation sign f in it precedes the elements x and y to which it is applied. There is another *infix* notation for algebraic operations, where the operation sign is placed between the elements x and y . Examples are the binary operations of addition and multiplication of numbers: $z = x + y$, $z = x \cdot y$. Sometimes special brackets play the role of the operation sign, while operands are separated by comma. The vector product of three-dimensional vectors yields an example of such notation: $\mathbf{z} = [\mathbf{x}, \mathbf{y}]$.

Let $\mathbb{K}$ be a numeric field. Under the numeric field in this book we shall understand one of three such fields: the field of rational numbers $\mathbb{K} = \mathbb{Q}$, the field of real numbers $\mathbb{K} = \mathbb{R}$, or the field of complex numbers $\mathbb{K} = \mathbb{C}$. The operation of

multiplication by numbers from the field $\mathbb{K}$ in a set M is a rule that maps each pair (α, x) consisting of a number $\alpha \in \mathbb{K}$ and of an element $x \in M$ to some element $y \in M$. The operation of multiplication by numbers is written in infix form: $y = \alpha \cdot x$. The multiplication sign in this notation is often omitted: $y = \alpha x$.

DEFINITION 2.1. A set V equipped with binary operation of addition and with the operation of multiplication by numbers from the field $\mathbb{K}$, is called a linear vector space over the field $\mathbb{K}$, if the following conditions are fulfilled:

- (1) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for all $\mathbf{u}, \mathbf{v} \in V$;
- (2) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$;
- (3) there is an element $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$; any such element is called a *zero element*;
- (4) for any $\mathbf{v} \in V$ and for any zero element $\mathbf{0}$ there is an element $\mathbf{v}' \in V$ such that $\mathbf{v} + \mathbf{v}' = \mathbf{0}$; it is called an *opposite element* for $\mathbf{v}$;
- (5) $\alpha \cdot (\mathbf{u} + \mathbf{v}) = \alpha \cdot \mathbf{u} + \alpha \cdot \mathbf{v}$ for any number $\alpha \in \mathbb{K}$ and for any two elements $\mathbf{u}, \mathbf{v} \in V$;
- (6) $(\alpha + \beta) \cdot \mathbf{v} = \alpha \cdot \mathbf{v} + \beta \cdot \mathbf{v}$ for any two numbers $\alpha, \beta \in \mathbb{K}$ and for any element $\mathbf{v} \in V$;
- (7) $\alpha \cdot (\beta \cdot \mathbf{v}) = (\alpha\beta) \cdot \mathbf{v}$ for any two numbers $\alpha, \beta \in \mathbb{K}$ and for any element $\mathbf{v} \in V$;
- (8) $1 \cdot \mathbf{v} = \mathbf{v}$ for the number $1 \in \mathbb{K}$ and for any element $\mathbf{v} \in V$.

The elements of a linear vector space are usually called the *vectors*, while the conditions (1)-(8) are called the *axioms of a linear vector space*. We shall distinguish *rational*, *real*, and *complex* linear vector spaces depending on which numeric field $\mathbb{K} = \mathbb{Q}$, $\mathbb{K} = \mathbb{R}$, or $\mathbb{K} = \mathbb{C}$ they are defined over. Most of the results in this book are valid for any numeric field $\mathbb{K}$. Formulating such results, we shall not specify the type of linear vector space.

Axioms (1) and (2) are the axiom of *commutativity*¹ and the axiom of *associativity* respectively. Axioms (5) and (6) express the *distributivity*.

THEOREM 2.1. *Algebraic operations in an arbitrary linear vector space V possess the following properties:*

- (9) zero vector $\mathbf{0} \in V$ is unique;
- (10) for any vector $\mathbf{v} \in V$ the vector $\mathbf{v}'$ opposite to $\mathbf{v}$ is unique;
- (11) the product of the number $0 \in \mathbb{K}$ and any vector $\mathbf{v} \in V$ is equal to zero vector: $0 \cdot \mathbf{v} = \mathbf{0}$;
- (12) the product of an arbitrary number $\alpha \in \mathbb{K}$ and zero vector is equal to zero vector: $\alpha \cdot \mathbf{0} = \mathbf{0}$;
- (13) the product of the number $-1 \in \mathbb{K}$ and the vector $\mathbf{v} \in V$ is equal to the opposite vector: $(-1) \cdot \mathbf{v} = \mathbf{v}'$.

PROOF. The properties (9)-(13) are immediate consequences of the axioms (1)-(8). Therefore, they are enumerated so that their numbers form successive series with the numbers of the axioms of a linear vector space.

Suppose that in a linear vector space there are two elements $\mathbf{0}$ and $\mathbf{0}'$ with the properties of zero vectors. Then for any vector $\mathbf{v} \in V$ due to the axiom (3) we

¹ The system of axioms (1)-(8) is excessive: the axiom (1) can be derived from other axioms. I am grateful to A. B. Muftakhov who communicated me this curious fact.

have $\mathbf{v} = \mathbf{v} + \mathbf{0}$ and $\mathbf{v} + \mathbf{0}' = \mathbf{v}$. Let's substitute $\mathbf{v} = \mathbf{0}'$ into the first equality and substitute $\mathbf{v} = \mathbf{0}$ into the second one. Taking into account the axiom (1), we get

$$\mathbf{0}' = \mathbf{0}' + \mathbf{0} = \mathbf{0} + \mathbf{0}' = \mathbf{0}.$$

This means that the vectors $\mathbf{0}$ and $\mathbf{0}'$ do actually coincide. The uniqueness of zero vector is proved.

Let $\mathbf{v}$ be some arbitrary vector in a vector space V . Suppose that there are two vectors $\mathbf{v}'$ and $\mathbf{v}''$ opposite to $\mathbf{v}$. Then

$$\mathbf{v} + \mathbf{v}' = \mathbf{0}, \quad \mathbf{v} + \mathbf{v}'' = \mathbf{0}.$$

The following calculations prove the uniqueness of opposite vector:

$$\begin{aligned} \mathbf{v}'' &= \mathbf{v}'' + \mathbf{0} = \mathbf{v}'' + (\mathbf{v} + \mathbf{v}') = (\mathbf{v}'' + \mathbf{v}) + \mathbf{v}' = \\ &= (\mathbf{v} + \mathbf{v}'') + \mathbf{v}' = \mathbf{0} + \mathbf{v}' = \mathbf{v}' + \mathbf{0} = \mathbf{v}'. \end{aligned}$$

In deriving $\mathbf{v}'' = \mathbf{v}'$ above we used the axiom (4), the associativity axiom (2) and we used twice the commutativity axiom (1).

Again, let $\mathbf{v}$ be some arbitrary vector in a vector space V . Let's take $\mathbf{x} = \mathbf{0} \cdot \mathbf{v}$, then let's add $\mathbf{x}$ with $\mathbf{x}$ and apply the distributivity axiom (6). As a result we get

$$\mathbf{x} + \mathbf{x} = \mathbf{0} \cdot \mathbf{v} + \mathbf{0} \cdot \mathbf{v} = (\mathbf{0} + \mathbf{0}) \cdot \mathbf{v} = \mathbf{0} \cdot \mathbf{v} = \mathbf{x}.$$

Thus we have proved that $\mathbf{x} + \mathbf{x} = \mathbf{x}$. Then we easily derive that $\mathbf{x} = \mathbf{0}$:

$$\mathbf{x} = \mathbf{x} + \mathbf{0} = \mathbf{x} + (\mathbf{x} + \mathbf{x}') = (\mathbf{x} + \mathbf{x}) + \mathbf{x}' = \mathbf{x} + \mathbf{x}' = \mathbf{0}.$$

Here we used the associativity axiom (2). The property (11) is proved.

Let α be some arbitrary number of a numeric field $\mathbb{K}$. Let's take $\mathbf{x} = \alpha \cdot \mathbf{0}$, where $\mathbf{0}$ is zero vector of a vector space V . Then

$$\mathbf{x} + \mathbf{x} = \alpha \cdot \mathbf{0} + \alpha \cdot \mathbf{0} = \alpha \cdot (\mathbf{0} + \mathbf{0}) = \alpha \cdot \mathbf{0} = \mathbf{x}.$$

Here we used the axiom (5) and the property of zero vector from the axiom (3). From the equality $\mathbf{x} + \mathbf{x} = \mathbf{x}$ it follows that $\mathbf{x} = \mathbf{0}$ (see above). Thus, the property (12) is proved.

Let $\mathbf{v}$ be some arbitrary vector of a vector space V . Let $\mathbf{x} = (-1) \cdot \mathbf{v}$. Applying axioms (8) and (6), for the vector $\mathbf{x}$ we derive

$$\mathbf{v} + \mathbf{x} = 1 \cdot \mathbf{v} + \mathbf{x} = 1 \cdot \mathbf{v} + (-1) \cdot \mathbf{v} = (1 + (-1)) \cdot \mathbf{v} = \mathbf{0} \cdot \mathbf{v} = \mathbf{0}.$$

The equality $\mathbf{v} + \mathbf{x} = \mathbf{0}$ just derived means that $\mathbf{x}$ is an opposite vector for the vector $\mathbf{v}$ in the sense of the axiom (4). Due to the uniqueness property (10) of the opposite vector we conclude that $\mathbf{x} = \mathbf{v}'$. Therefore, $(-1) \cdot \mathbf{v} = \mathbf{v}'$. The theorem is completely proved. $\square$

Due to the commutativity and associativity axioms we need not worry about setting brackets and about the order of the summands when writing the sums of vectors. The property (13) and the axioms (7) and (8) yield

$$(-1) \cdot \mathbf{v}' = (-1) \cdot ((-1) \cdot \mathbf{v}) = ((-1)(-1)) \cdot \mathbf{v} = 1 \cdot \mathbf{v} = \mathbf{v}.$$

This equality shows that the notation $\mathbf{v}' = -\mathbf{v}$ for an opposite vector is quite natural. In addition, we can write

$$-\alpha \cdot \mathbf{v} = -(\alpha \cdot \mathbf{v}) = (-1) \cdot (\alpha \cdot \mathbf{v}) = (-\alpha) \cdot \mathbf{v}.$$

The operation of *subtraction* is an opposite operation for the vector addition. It is determined as the addition with the opposite vector: $\mathbf{x} - \mathbf{y} = \mathbf{x} + (-\mathbf{y})$. The following properties of the operation of vector subtraction

$$\begin{aligned} (\mathbf{a} + \mathbf{b}) - \mathbf{c} &= \mathbf{a} + (\mathbf{b} - \mathbf{c}), \\ (\mathbf{a} - \mathbf{b}) + \mathbf{c} &= \mathbf{a} - (\mathbf{b} - \mathbf{c}), \\ (\mathbf{a} - \mathbf{b}) - \mathbf{c} &= \mathbf{a} - (\mathbf{b} + \mathbf{c}), \\ \alpha \cdot (\mathbf{x} - \mathbf{y}) &= \alpha \cdot \mathbf{x} - \alpha \cdot \mathbf{y} \end{aligned}$$

make the calculations with vectors very simple and quite similar to the calculations with numbers. Proof of the above properties is left to the reader.

Let's consider some examples of linear vector spaces. Real *arithmetic vector space* $\mathbb{R}^n$ is determined as a set of ordered n -tuples of real numbers $x^1, \dots, x^n$. Such n -tuples are represented in the form of *column vectors*. Algebraic operations with column vectors are determined as the operations with their components:

$$\begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix} + \begin{pmatrix} y^1 \\ y^2 \\ \vdots \\ y^n \end{pmatrix} = \begin{pmatrix} x^1 + y^1 \\ x^2 + y^2 \\ \vdots \\ x^n + y^n \end{pmatrix} \quad \alpha \cdot \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix} = \begin{pmatrix} \alpha \cdot x^1 \\ \alpha \cdot x^2 \\ \vdots \\ \alpha \cdot x^n \end{pmatrix} \quad (2.1)$$

We leave to the reader to check the fact that the set $\mathbb{R}^n$ of all ordered n -tuples with algebraic operations (2.1) is a linear vector space over the field $\mathbb{R}$ of real numbers. Rational arithmetic vector space $\mathbb{Q}^n$ over the field $\mathbb{Q}$ of rational numbers and complex arithmetic vector space $\mathbb{C}^n$ over the field $\mathbb{C}$ of complex numbers are defined in a similar way.

Let's consider the set of m -times continuously differentiable real-valued functions on the segment $[-1, 1]$ of real axis. This set is usually denoted as $C^m([-1, 1])$. The operations of addition and multiplication by numbers in $C^m([-1, 1])$ are defined as pointwise operations. This means that the value of the function $f + g$ at a point a is the sum of the values of f and g at that point. In a similar way, the value of the function $\alpha \cdot f$ at the point a is the product of two numbers α and $f(a)$. It is easy to verify that the set of functions $C^m([-1, 1])$ with pointwise algebraic operations of addition and multiplication by numbers is a linear vector space over the field of real numbers $\mathbb{R}$. The reader can easily verify this fact.

DEFINITION 2.2. A non-empty subset $U \subset V$ in a linear vector space V over a numeric field $\mathbb{K}$ is called a subspace of the space V if:

- (1) from $\mathbf{u}_1, \mathbf{u}_2 \in U$ it follows that $\mathbf{u}_1 + \mathbf{u}_2 \in U$;
- (2) from $\mathbf{u} \in U$ it follows that $\alpha \cdot \mathbf{u} \in U$ for any number $\alpha \in \mathbb{K}$.

Let U be a subspace of a linear vector space V . Let's regard U as an isolated set. Due to the above conditions (1) and (2) this set is closed with respect to operations of addition and multiplication by numbers. It is easy to show that

zero vector is an element of U and for any $\mathbf{u} \in U$ the opposite vector $\mathbf{u}'$ also is an element of U . These facts follow from $\mathbf{0} = 0 \cdot \mathbf{u}$ and $\mathbf{u}' = (-1) \cdot \mathbf{u}$. Relying upon these facts one can easily prove that any subspace $U \subset V$, when considered as an isolated set, is a linear vector space over the field $\mathbb{K}$. Indeed, we have already shown that axioms (3) and (4) are valid for it. Verifying axioms (1), (2) and remaining axioms (5)-(8) consists in checking equalities written in terms of the operations of addition and multiplication by numbers. Being fulfilled for arbitrary vectors of V , these equalities are obviously fulfilled for vectors of subset $U \subset V$. Since U is closed with respect to algebraic operations, it makes sure that all calculations in these equalities are performed within the subset U .

As the examples of the concept of subspace we can mention the following subspaces in the functional space $C^m([-1, 1])$:

- the subspace of even functions ($f(-x) = f(x)$);
- the subspace of odd functions ($f(-x) = -f(x)$);
- the subspace of polynomials ($f(x) = a_n x^n + \dots + a_1 x + a_0$).

§ 3. Linear dependence and linear independence.

Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be a system of vectors some from some linear vector space V . Applying the operations of multiplication by numbers and addition to them we can produce the following expressions with these vectors:

$$\mathbf{v} = \alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_n \cdot \mathbf{v}_n. \quad (3.1)$$

An expression of the form (3.1) is called a *linear combination* of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$. The numbers $\alpha_1, \dots, \alpha_n$ are taken from the field $\mathbb{K}$; they are called the *coefficients* of the linear combination (3.1), while vector $\mathbf{v}$ is called the *value* of this linear combination. Linear combination is said to be *zero* or *equal to zero* if its value is zero.

A linear combination is called *trivial* if all its coefficients are equal to zero: $\alpha_1 = \dots = \alpha_n = 0$. Otherwise it is called *nontrivial*.

DEFINITION 3.1. A system of vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ in linear vector space V is called *linearly dependent* if there exists some nontrivial linear combination of these vectors equal to zero.

DEFINITION 3.2. A system of vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ in linear vector space V is called *linearly independent* if any linear combination of these vectors being equal to zero is necessarily trivial.

The concept of linear independence is obtained by direct logical negation of the concept of linear dependence. The reader can give several equivalent statements defining this concept. Here we give only one of such statements which, to our knowledge, is most convenient in what follows.

Let's introduce one more concept related to linear combinations. We say that vector $\mathbf{v}$ is *linearly expressed* through the vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ if $\mathbf{v}$ is the value of some linear combination composed of $\mathbf{v}_1, \dots, \mathbf{v}_n$.

THEOREM 3.1. *The relation of linear dependence of vectors in a linear vector space has the following basic properties:*

- (1) *any system of vectors comprising zero vector is linearly dependent;*
- (2) *any system of vectors comprising linearly dependent subsystem is linearly dependent in whole;*
- (3) *if a system of vectors is linearly dependent, then at least one of these vectors is linearly expressed through others;*
- (4) *if a system of vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ is linearly independent and if adding the next vector $\mathbf{v}_{n+1}$ to it we make it linearly dependent, then the vector $\mathbf{v}_{n+1}$ is linearly expressed through previous vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$;*
- (5) *if a vector $\mathbf{x}$ is linearly expressed through the vectors $\mathbf{y}_1, \dots, \mathbf{y}_m$ and if each one of the vectors $\mathbf{y}_1, \dots, \mathbf{y}_m$ is linearly expressed through $\mathbf{z}_1, \dots, \mathbf{z}_n$, then $\mathbf{x}$ is linearly expressed through $\mathbf{z}_1, \dots, \mathbf{z}_n$.*

PROOF. Suppose that a system of vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ comprises zero vector. For the sake of certainty we can assume that $\mathbf{v}_k = \mathbf{0}$. Let's compose the following linear combination of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$:

$$0 \cdot \mathbf{v}_1 + \dots + 0 \cdot \mathbf{v}_{k-1} + 1 \cdot \mathbf{v}_k + 0 \cdot \mathbf{v}_{k+1} + \dots + 0 \cdot \mathbf{v}_n = \mathbf{0}.$$

This linear combination is nontrivial since the coefficient of vector $\mathbf{v}_k$ is nonzero. And its value is equal to zero. Hence, the vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly dependent. The property (1) is proved. Suppose that a system of vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ comprises a linear dependent subsystem. Since linear dependence is not sensible to the order in which the vectors in a system are enumerated, we can assume that first k vectors form linear dependent subsystem in it. Then there exists some nontrivial linear combination of these k vectors being equal to zero:

$$\alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_k \cdot \mathbf{v}_k = \mathbf{0}.$$

Let's expand this linear combination by adding other vectors with zero coefficients:

$$\alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_k \cdot \mathbf{v}_k + 0 \cdot \mathbf{v}_{k+1} + \dots + 0 \cdot \mathbf{v}_n = \mathbf{0}.$$

It is obvious that the resulting linear combination is nontrivial and its value is equal to zero. Hence, the vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly dependent. The property (2) is proved.

Let assume that the vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly dependent. Then there exists a nontrivial linear combination of them being equal to zero:

$$\alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_n \cdot \mathbf{v}_n = \mathbf{0}. \quad (3.2)$$

Non-triviality of the linear combination (3.2) means that at least one of its coefficients is nonzero. Suppose that $\alpha_k \neq 0$. Let's write (3.2) in more details:

$$\alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_k \cdot \mathbf{v}_k + \dots + \alpha_n \cdot \mathbf{v}_n = \mathbf{0}.$$

Let's move the term $\alpha_k \cdot \mathbf{v}_k$ to the right hand side of the above equality, and then let's divide the equality by $-\alpha_k$:

$$\mathbf{v}_k = -\frac{\alpha_1}{\alpha_k} \cdot \mathbf{v}_1 - \dots - \frac{\alpha_{k-1}}{\alpha_k} \cdot \mathbf{v}_{k-1} - \frac{\alpha_{k+1}}{\alpha_k} \cdot \mathbf{v}_{k+1} - \dots - \frac{\alpha_n}{\alpha_k} \cdot \mathbf{v}_n.$$

Now we see that the vector $\mathbf{v}_k$ is linearly expressed through other vectors of the system. The property (3) is proved.

Let's consider a linearly independent system of vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ such that adding the next vector $\mathbf{v}_{n+1}$ to it we make it linearly dependent. Then there is some nontrivial linear combination of vectors $\mathbf{v}_1, \dots, \mathbf{v}_{n+1}$ being equal to zero:

$$\alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_n \cdot \mathbf{v}_n + \alpha_{n+1} \cdot \mathbf{v}_{n+1} = \mathbf{0}.$$

Let's prove that $\alpha_{n+1} \neq 0$. If, conversely, we assume that $\alpha_{n+1} = 0$, we would get the nontrivial linear combination of n vectors being equal to zero:

$$\alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_n \cdot \mathbf{v}_n = \mathbf{0}.$$

This contradicts to the linear independence of the first n vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$. Hence, $\alpha_{n+1} \neq 0$, and we can apply the trick already used above:

$$\mathbf{v}_{n+1} = -\frac{\alpha_1}{\alpha_{n+1}} \cdot \mathbf{v}_1 - \dots - \frac{\alpha_n}{\alpha_{n+1}} \cdot \mathbf{v}_n.$$

This expression for the vector $\mathbf{v}_{n+1}$ completes the proof of the property (4).

Suppose that the vector $\mathbf{x}$ is linearly expressed through $\mathbf{y}_1, \dots, \mathbf{y}_m$, and each one of the vectors $\mathbf{y}_1, \dots, \mathbf{y}_m$ is linearly expressed through $\mathbf{z}_1, \dots, \mathbf{z}_n$. This fact is expressed by the following formulas:

$$x = \sum_{i=1}^m \alpha_i \cdot y_i, \quad y_i = \sum_{j=1}^n \beta_{ij} \cdot z_j.$$

Substituting second formula into the first one, for the vector $\mathbf{x}$ we get

$$x = \sum_{i=1}^m \alpha_i \cdot \left(\sum_{j=1}^n \beta_{ij} \cdot z_j \right) = \sum_{j=1}^n \left(\sum_{i=1}^m \alpha_i \beta_{ij} \right) \cdot z_j$$

The above expression for the vector $\mathbf{x}$ shows that it is linearly expressed through vectors $\mathbf{z}_1, \dots, \mathbf{z}_n$. The property (5) is proved. This completes the proof of theorem 3.1 in whole. $\square$

Note the following important consequence that follows from the property (2) in the theorem 3.1.

COROLLARY. *Any subsystem in a linearly independent system of vectors is linearly independent.*

THEOREM 3.2 (STEINITZ). *If the vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ are linear independent and if each of them is expressed through the vectors $\mathbf{y}_1, \dots, \mathbf{y}_m$, then $m \geq n$.*

Suppose that the theorem holds for the case $n = k$. Under this assumption let's prove that it is valid for $n = k + 1$. If $n = k + 1$ we have a system of linearly independent vectors $\mathbf{x}_1, \dots, \mathbf{x}_{k+1}$, each vector being expressed through the vectors of another system $\mathbf{y}_1, \dots, \mathbf{y}_m$. We express this fact by formulas

[illegible]

$$\mathbf{x}_{k+1} = \beta_1 \cdot \mathbf{y}_1 + \dots + \beta_m \cdot \mathbf{y}_m.$$
$$\mathbf{y}_m = \frac{1}{\beta_m} \cdot \mathbf{x}_{k+1} - \frac{\beta_1}{\beta_m} \cdot \mathbf{y}_1 - \dots - \frac{\beta_{m-1}}{\beta_m} \cdot \mathbf{y}_{m-1}. \quad (3.4)$$
$$\mathbf{x}_i - \frac{\alpha_{im}}{\beta_m} \cdot \mathbf{x}_{k+1} = \sum_{j=1}^{m-1} \left(\alpha_{ij} - \beta_j \frac{\alpha_{im}}{\beta_m} \right) \cdot \mathbf{y}_j, \quad (3.5)$$
$$x_i^* = x_i - \frac{\alpha_{im}}{\beta_m} \cdot x_{k+1}, \quad \alpha_{ij}^* = \alpha_{ij} - \beta_j \frac{\alpha_{im}}{\beta_m}. \quad (3.6)$$
[illegible]

According to the above formulas, k vectors $\mathbf{x}_1^*, \dots, \mathbf{x}_k^*$ are linearly expressed through $\mathbf{y}_1, \dots, \mathbf{y}_{m-1}$. In order to apply the inductive hypothesis we need to show that the vectors $\mathbf{x}_1^*, \dots, \mathbf{x}_k^*$ are linearly independent. Let's consider a linear combination of these vectors being equal to zero:

$$\gamma_1 \cdot \mathbf{x}_1^* + \dots + \gamma_k \cdot \mathbf{x}_k^* = \mathbf{0}. \quad (3.8)$$

Substituting (3.6) for x_i^* in (3.8), upon collecting similar terms, we get

$$\gamma_1 \cdot \mathbf{x}_1 + \dots + \gamma_k \cdot \mathbf{x}_k - \left(\sum_{i=1}^k \gamma_i \frac{\alpha_{im}}{\beta_m} \right) \cdot \mathbf{x}_{k+1} = \mathbf{0}.$$

Due to the linear independence of the initial system of vectors $\mathbf{x}_1, \dots, \mathbf{x}_{k+1}$ we derive $\gamma_1 = \dots = \gamma_k = 0$. Hence, the linear combination (3.8) is trivial, which proves the linear independence of vectors $\mathbf{x}_1^*, \dots, \mathbf{x}_k^*$. Now, applying the inductive hypothesis to the relationships (3.7), we get $m-1 \geq k$. The required inequality $m \geq k+1$ proving the theorem for the case $n = k+1$ is an immediate consequence of $m \geq k+1$. So, the inductive step is completed and the theorem is proved. $\square$

§ 4. Spanning systems and bases.

Let $S \subset V$ be some non-empty subset in a linear vector space V . The set S can consist of either finite number of vectors, or of infinite number of vectors. We denote by $\langle S \rangle$ the set of all vectors, each of which is linearly expressed through some finite number of vectors taken from S :

$$\langle S \rangle = \{ \mathbf{v} \in V : \exists n (\mathbf{v} = \alpha_1 \cdot \mathbf{s}_1 + \dots + \alpha_n \cdot \mathbf{s}_n, \text{ where } \mathbf{s}_i \in S) \}.$$

This set $\langle S \rangle$ is called the *linear span* of a subset $S \subset V$.

THEOREM 4.1. *The linear span of any subset $S \subset V$ is a subspace in a linear vector space V .*

PROOF. In order to prove this theorem it is sufficient to check two conditions from the definition 2.2 for $\langle S \rangle$. Suppose that $\mathbf{u}_1, \mathbf{u}_2 \in \langle S \rangle$. Then

$$\begin{aligned} \mathbf{u}_1 &= \alpha_1 \cdot \mathbf{s}_1 + \dots + \alpha_n \cdot \mathbf{s}_n, \\ \mathbf{u}_2 &= \beta_1 \cdot \mathbf{s}_1^* + \dots + \beta_m \cdot \mathbf{s}_m^*. \end{aligned}$$

Adding these two equalities, we see that the vector $\mathbf{u}_1 + \mathbf{u}_2$ also is expressed as a linear combination of some finite number of vectors taken from S . Therefore, we have $\mathbf{u}_1 + \mathbf{u}_2 \in \langle S \rangle$.

Now suppose that $\mathbf{u} \in \langle S \rangle$. Then $\mathbf{u} = \alpha_1 \cdot \mathbf{s}_1 + \dots + \alpha_n \cdot \mathbf{s}_n$. For the vector $\alpha \cdot \mathbf{u}$, from this equality we derive

$$\alpha \cdot \mathbf{u} = (\alpha \alpha_1) \cdot \mathbf{s}_1 + \dots + (\alpha \alpha_n) \cdot \mathbf{s}_n.$$

Hence, $\alpha \cdot \mathbf{u} \in \langle S \rangle$. Both conditions (1) and (2) from the definition 2.2 for $\langle S \rangle$ are fulfilled. Thus, the theorem is proved. $\square$

THEOREM 4.2. *The operation of passing to the linear span in a linear vector space V possesses the following properties:*

- (1) *if $S \subset U$ and if U is a subspace in V , then $\langle S \rangle \subset U$;*
- (2) *the linear span of a subset $S \subset V$ is the intersection of all subspaces comprising this subset S .*

PROOF. Let $\mathbf{u} \in \langle S \rangle$ and $S \subset U$, where U is a subspace. Then for the vector $\mathbf{u}$ we have $\mathbf{u} = \alpha_1 \cdot \mathbf{s}_1 + \dots + \alpha_n \cdot \mathbf{s}_n$, where $\mathbf{s}_i \in S$. But $\mathbf{s}_i \in S$ and $S \subset U$ implies $\mathbf{s}_i \in U$. Since U is a subspace, the value of any linear combination of its elements again is an element of U . Hence, $\mathbf{u} \in U$. This proves the inclusion $\langle S \rangle \subset U$.

Let's denote by W the intersection of all subspaces of V comprising the subset S . Due to the property (1), which is already proved, the subset $\langle S \rangle$ is included into each of such subspaces. Therefore, $\langle S \rangle \subset W$. On the other hand, $\langle S \rangle$ is a subspace of V comprising the subset S (see theorem 4.1). Hence, $\langle S \rangle$ is among those subspaces forming W . Then $W \subset \langle S \rangle$. From the two inclusions $\langle S \rangle \subset W$ and $W \subset \langle S \rangle$ it follows that $\langle S \rangle = W$. The theorem is proved. $\square$

Let $\langle S \rangle = U$. Then we say that the subset $S \subset V$ *spans* the subspace U , i. e. S *generates* U by means of the linear combinations. This terminology is supported by the following definition.

DEFINITION 4.1. A subset $S \subset V$ is called a *generating subset* or a *spanning system of vectors* in a linear vector space V if $\langle S \rangle = V$.

A linear vector space V can have multiple spanning systems. Therefore the problem of choosing of a minimal (in some sense) spanning system is reasonable.

DEFINITION 4.2. A spanning system of vectors $S \subset V$ in a linear vector space V is called a *minimal spanning system* if none of smaller subsystems $S' \subsetneq S$ is a spanning system in V , i. e. if $\langle S' \rangle \neq V$ for all $S' \subsetneq S$.

DEFINITION 4.3. A system of vectors $S \subset V$ is called *linearly independent* if any finite subsystem of vectors $\mathbf{s}_1, \dots, \mathbf{s}_n$ taken from S is linearly independent.

This definition extends the definition 3.2 for the case of infinite systems of vectors. As for the spanning systems, the relation of the properties of minimality and linear independence for them is determined by the following theorem.

THEOREM 4.3. *A spanning system of vectors $S \subset V$ is minimal if and only if it is linearly independent.*

PROOF. If a spanning system of vectors $S \subset V$ is linearly dependent, then it contains some finite linearly dependent set of vectors $\mathbf{s}_1, \dots, \mathbf{s}_n$. Due to the item (3) in the statement of theorem 3.1 one of these vectors $\mathbf{s}_k$ is linearly expressed through others. Then the subsystem $S' = S \setminus \{\mathbf{s}_k\}$ obtained by omitting this vector $\mathbf{s}_k$ from S is a spanning system in V . This fact obviously contradicts the minimality of S (see definition 4.2 above). Therefore any minimal spanning system of vectors in V is linearly independent.

If a spanning system of vectors $S \subset V$ is not minimal, then there is some smaller spanning subsystem $S' \subsetneq S$, i. e. subsystem S' such that

$$\langle S' \rangle = \langle S \rangle = V. \quad (4.1)$$

In this case we can choose some vector $\mathbf{s}_0 \in S$ such that $\mathbf{s}_0 \notin S'$. Due to (4.1) this vector is an element of $\langle S' \rangle$. Hence, $\mathbf{s}_0$ is linearly expressed through some finite number of vectors taken from the subsystem S' :

$$\mathbf{s}_0 = \alpha_1 \cdot \mathbf{s}_1 + \dots + \alpha_n \cdot \mathbf{s}_n. \quad (4.2)$$

One can easily transform (4.2) to the form of a linear combination equal to zero:

$$(-1) \cdot \mathbf{s}_0 + \alpha_1 \cdot \mathbf{s}_1 + \dots + \alpha_n \cdot \mathbf{s}_n = \mathbf{0}. \quad (4.3)$$

This linear combination is obviously nontrivial. Thus, we have found that the vectors $\mathbf{s}_0, \dots, \mathbf{s}_n$ form a finite linearly dependent subset of S . Hence, S is linearly dependent (see the item (2) in theorem 3.1 and the definition 4.2). This fact means that any linearly independent spanning system of vector in V is minimal. $\square$

DEFINITION 4.4. A linear vector space V is called *finite dimensional* if there is some finite spanning system of vectors $S = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ in it.

In an arbitrary linear vector space V there is at least one spanning system, e. g. $S = V$. However, the problem of existence of minimal spanning systems in general case is nontrivial. The solution of this problem is positive, but it is not elementary and it is not constructive. This problem is solved with the use of the *axiom of choice* (see [1]). Finite dimensional vector spaces are distinguished due to the fact that the proof of existence of minimal spanning systems for them is elementary.

THEOREM 4.4. In a finite dimensional linear vector space V there is at least one minimal spanning system of vectors. Any two of such systems $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ and $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ have the same number of elements n . This number n is called the *dimension* of V , it is denoted as $n = \dim V$.

PROOF. Let $S = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ be some finite spanning system of vectors in a finite-dimensional linear vector space V . If this system is not minimal, then it is linear dependent. Hence, one of its vectors is linearly expressed through others. This vector can be omitted and we get the smaller spanning system S' consisting of $k-1$ vectors. If S' is not minimal again, then we can iterate the process getting one less vectors in each step. Ultimately, we shall get a minimal spanning system $S_{\min}$ in V with finite number of vectors n in it:

$$S_{\min} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}. \quad (4.4)$$

Usually, the minimal spanning system of vectors (4.4) is not unique. Suppose that $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ is some other minimal spanning system in V . Both systems $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ and $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ are linearly independent and

$$\begin{aligned} x^i &\in \langle \mathbf{y}_1, \dots, \mathbf{y}_n \rangle \text{ for } i = 1, \dots, m, \\ y^i &\in \langle \mathbf{x}_1, \dots, \mathbf{x}_m \rangle \text{ for } i = 1, \dots, n. \end{aligned} \quad (4.5)$$

Due to (4.5) we can apply Steinitz theorem 3.2 to the systems of vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ and $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$. As a result we get two inequalities $n \geq m$ and $m \geq n$. Therefore, $m = n = \dim V$. The theorem is proved. $\square$

The dimension $\dim V$ is an integer invariant of a finite-dimensional linear vector space. If $\dim V = n$, then such a space is called an n -dimensional space. Returning to the examples of linear vector spaces considered in § 2, note that $\dim \mathbb{R}^n = n$, while the functional space $C^m([-1, 1])$ is not finite-dimensional at all.

THEOREM 4.5. *Let V be a finite dimensional linear vector space. Then the following propositions are valid:*

- (1) *the number of vectors in any linearly independent system of vectors $\mathbf{x}_1, \dots, \mathbf{x}_k$ in V is not greater than the dimension of V ;*
- (2) *any subspace U of the space V is finite-dimensional and $\dim U \leq \dim V$;*
- (3) *for any subspace U in V if $\dim U = \dim V$, then $U = V$;*
- (4) *any linearly independent system of n vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$, where $n = \dim V$, is a spanning system in V .*

PROOF. Suppose that $\dim V = n$. Let's fix some minimal spanning system of vectors $\mathbf{y}_1, \dots, \mathbf{y}_n$ in V . Then each vector of the linear independent system of vectors $\mathbf{x}_1, \dots, \mathbf{x}_k$ in proposition (1) is linearly expressed through $\mathbf{y}_1, \dots, \mathbf{y}_n$. Applying Steinitz theorem 3.2, we get the inequality $k \leq n$. The first proposition of theorem is proved.

Let's consider all possible linear independent systems $\mathbf{u}_1, \dots, \mathbf{u}_k$ composed by the vectors of a subspace U . Due to the proposition (1), which is already proved, the number of vectors in such systems is restricted. It is not greater than $n = \dim V$. Therefore we can assume that $\mathbf{u}_1, \dots, \mathbf{u}_k$ is a linearly independent system with maximal number of vectors: $k = k_{\max} \leq n = \dim V$. If $\mathbf{u}$ is an arbitrary vector of the subspace U and if we add it to the system $\mathbf{u}_1, \dots, \mathbf{u}_k$, we get a linearly dependent system; this is because $k = k_{\max}$. Now, applying the property (4) from the theorem 3.1, we conclude that the vector $\mathbf{u}$ is linearly expressed through the vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$. Hence, the vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ form a finite spanning system in U . It is minimal since it is linearly independent (see theorem 4.3). Finite dimensionality of U is proved. The estimate for its dimension follows from the above inequality: $\dim U = k \leq n = \dim V$.

Let U again be a subspace in V . Assume that $\dim U = \dim V = n$. Let's choose some minimal spanning system of vectors $\mathbf{u}_1, \dots, \mathbf{u}_n$ in U . It is linearly independent. Adding an arbitrary vector $\mathbf{v} \in V$ to this system, we make it linearly dependent since in V there is no linearly independent system with $(n + 1)$ vectors (see proposition (1), which is already proved). Furthermore, applying the property (3) from the theorem 3.1 to the system $\mathbf{u}_1, \dots, \mathbf{u}_n, \mathbf{v}$, we find that

$$\mathbf{v} = \alpha_1 \cdot \mathbf{u}_1 + \dots + \alpha_m \cdot \mathbf{u}_m.$$

This formula means that $\mathbf{v} \in U$, where $\mathbf{v}$ is an arbitrary vector of the space V . Therefore, $U = V$. The third proposition of the theorem is proved.

Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be a linearly independent system of n vectors in V , where n is equal to the dimension of the space V . Denote by U the linear span of this system of vectors: $U = \langle \mathbf{x}_1, \dots, \mathbf{x}_n \rangle$. Since $\mathbf{x}_1, \dots, \mathbf{x}_n$ are linearly independent, they form a minimal spanning system in U . Therefore, $\dim U = n = \dim V$. Now, applying proposition (3) of the theorem, we get

$$\langle \mathbf{x}_1, \dots, \mathbf{x}_n \rangle = U = V.$$

This equality proves the fourth proposition of theorem 4.5 and completes the proof of the theorem in whole. $\square$

DEFINITION 4.5. A minimal spanning system $\mathbf{e}_1, \dots, \mathbf{e}_n$ with some fixed order of vectors in it is called a *basis* of a finite-dimensional vector space V .

THEOREM (BASIS CRITERION). *An ordered system of vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ is a basis in a finite-dimensional vector space V if and only if*

- (1) *the vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ are linearly independent;*
- (2) *an arbitrary vector of the space V is linearly expressed through $\mathbf{e}_1, \dots, \mathbf{e}_n$.*

Proof is obvious. The second condition of theorem means that the vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ form a spanning system in V , while the first condition is equivalent to its minimality.

In essential, theorem 4.6 simply reformulates the definition 4.5. We give it here in order to simplify the terminology. The terms «spanning system» and «minimal spanning system» are huge and inconvenient for often usage.

THEOREM 4.7. *Let $\mathbf{e}_1, \dots, \mathbf{e}_s$ be a basis in a subspace $U \subset V$ and let $\mathbf{v} \in V$ be some vector outside this subspace: $\mathbf{v} \notin U$. Then the system of vectors $\mathbf{e}_1, \dots, \mathbf{e}_s, \mathbf{v}$ is a linearly independent system.*

PROOF. Indeed, if the system of vectors $\mathbf{e}_1, \dots, \mathbf{e}_s, \mathbf{v}$ is linearly dependent, while $\mathbf{e}_1, \dots, \mathbf{e}_s$ is a linearly independent system, then $\mathbf{v}$ is linearly expressed through the vectors $\mathbf{e}_1, \dots, \mathbf{e}_s$, thus contradicting the condition $\mathbf{v} \notin U$. This contradiction proves the theorem 4.7. $\square$

THEOREM 4.8 (ON COMPLETING THE BASIS). *Let U be a subspace in a finite-dimensional linear vector space V . Then any basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ of U can be completed up to a basis $\mathbf{e}_1, \dots, \mathbf{e}_s, \mathbf{e}_{s+1}, \dots, \mathbf{e}_n$ in V .*

PROOF. Let's denote $U = U_0$. If $U_0 = V$, then there is no need to complete the basis since $\mathbf{e}_1, \dots, \mathbf{e}_s$ is a basis in V . Otherwise, if $U_0 \neq V$, then let's denote by $\mathbf{e}_{s+1}$ some arbitrary vector of V taken outside the subspace U_0 . According to the above theorem 4.7, the vectors $\mathbf{e}_1, \dots, \mathbf{e}_s, \mathbf{e}_{s+1}$ are linearly independent.

Let's denote by U_1 the linear span of vectors $\mathbf{e}_1, \dots, \mathbf{e}_s, \mathbf{e}_{s+1}$. For the subspace U_1 we have the same two mutually exclusive options $U_1 = V$ or $U_1 \neq V$, as we previously had for the subspace U_0 . If $U_1 = V$, then the process of completing the basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ is over. Otherwise, we can iterate the process and get a chain of subspaces enclosed into each other:

$$U_0 \subsetneq U_1 \subsetneq U_2 \subsetneq \dots$$

This chain of subspaces cannot be infinite since the dimension of every next subspace is one as greater than the dimension of previous subspace, and the dimensions of all subspaces are not greater than the dimension of V . The process of completing the basis will be finished in $(n - s)$ -th step, where $U_{n-s} = V$. $\square$

§ 5. Coordinates. Transformation of the coordinates of a vector under a change of basis.

Let V be some finite-dimensional linear vector space over the field $\mathbb{K}$ and let $\dim V = n$. In this section we shall consider only finite-dimensional spaces. Let's

choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in V . Then an arbitrary vector $x \in V$ can be expressed as linear combination of the basis vectors:

$$\mathbf{x} = x^1 \cdot \mathbf{e}_1 + \dots + x^n \cdot \mathbf{e}_n. \quad (5.1)$$

The linear combination (5.1) is called the *expansion* of the vector $\mathbf{x}$ in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. Its coefficients $x^1, \dots, x^n$ are the elements of the numeric field $\mathbb{K}$. They are called the *components* or the *coordinates* of the vector $\mathbf{x}$ in this basis.

We use upper indices for the literal notations of the coordinates of a vector $\mathbf{x}$ in (5.1). The usage of upper indices for the coordinates of vectors is determined by special convention, which is known as *tensorial notation*. It was introduced for to simplify huge calculations in differential geometry and in theory of relativity (see [2] and [3]). Other rules of tensorial notation are discussed in coordinate theory of tensors (see [7]¹).

THEOREM 5.1. *For any vector $\mathbf{x} \in V$ its expansion in a basis of a linear vector space V is unique.*

PROOF. The existence of an expansion (5.1) for a vector $\mathbf{x}$ follows from the item (2) of theorem 4.7. Assume that there is another expansion

$$\mathbf{x} = x'^1 \cdot \mathbf{e}_1 + \dots + x'^n \cdot \mathbf{e}_n. \quad (5.2)$$

Subtracting (5.1) from this equality, we get

$$\mathbf{x} = (x'^1 - x^1) \cdot \mathbf{e}_1 + \dots + (x'^n - x^n) \cdot \mathbf{e}_n. \quad (5.3)$$

Since basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ are linearly independent, from the equality (5.3) it follows that the linear combination (5.3) is trivial: $x'^i - x^i = 0$. Then

$$x'^1 = x^1, \dots, x'^n = x^n.$$

Hence the expansions (5.1) and (5.2) do coincide. The uniqueness of the expansion (5.1) is proved. $\square$

Having chosen some basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in a space V and expanding a vector $\mathbf{x}$ in this base we can write its coordinates in the form of column vectors. Due to the theorem 5.1 this determines a bijective map $\psi : V \rightarrow \mathbb{K}^n$. It is easy to verify that

$$\psi(x + y) = \begin{pmatrix} x^1 + y^1 \\ \vdots \\ x^n + y^n \end{pmatrix}, \quad \psi(\alpha \cdot x) = \begin{pmatrix} \alpha \cdot x^1 \\ \vdots \\ \alpha \cdot x^n \end{pmatrix}. \quad (5.4)$$

The above formulas (5.4) show that a basis is a very convenient tool when dealing with vectors. In a basis algebraic operations with vectors are replaced by algebraic operations with their coordinates, i. e. with numeric quantities. However, coordinate approach has one disadvantage. The mapping ψ essentially depends on the basis we choose. And there is no canonic choice of basis. In general, none of basis is preferable with respect to another. Therefore we should be ready to

¹ The reference [7] is added in 2004 to English translation of this book.

consider various bases and should be able to recalculate the coordinates of vectors when passing from a basis to another basis.

Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ be two arbitrary bases in a linear vector space V . We shall call them «wavy» basis and «non-wavy» basis (because of tilde sign we use for denoting the vectors of one of them). The non-wavy basis will also be called the *initial basis* or the *old basis*, and the wavy one will be called the *new basis*. Taking i -th vector of new (wavy) basis, we expand it in the old basis:

$$\tilde{\mathbf{e}}_i = S_i^1 \cdot \mathbf{e}_1 + \dots + S_i^n \cdot \mathbf{e}_n. \quad (5.5)$$

According to the tensorial notation, the coordinates of the vector $\tilde{\mathbf{e}}_i$ in the expansion (5.5) are specified by upper index. The lower index i specifies the number of the vector $\tilde{\mathbf{e}}_i$ being expanded. Totally in the expansion (5.5) we determine n^2 numbers; they are usually arranged into a matrix:

$$S = \begin{pmatrix} S_1^1 & \dots & S_n^1 \\ \vdots & \ddots & \vdots \\ S_1^n & \dots & S_n^n \end{pmatrix}. \quad (5.6)$$

Upper index j of the matrix element S_i^j specifies the row number; lower index i specifies the column number. The matrix S in (5.6) the *direct transition matrix* for passing from the old basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ to the new basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$.

Swapping the bases $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ we can write the expansion of the vector $\mathbf{e}_j$ in wavy basis:

$$\mathbf{e}_j = T_j^1 \cdot \tilde{\mathbf{e}}_1 + \dots + T_j^n \cdot \tilde{\mathbf{e}}_n. \quad (5.7)$$

The coefficients of the expansion (5.7) determine the matrix T , which is called the *inverse transition matrix*. Certainly, the usage of terms «direct» and «inverse» here is relative; it depends on which basis is considered as an old basis and which one is taken for a new one.

THEOREM 5.2. *The direct transition matrix S and the inverse transition matrix T determined by the expansions (5.5) and (5.7) are inverse to each other.*

Remember that two square matrices are inverse to each other if their product is equal to unit matrix: $ST = 1$. Here we do not define the matrix multiplication assuming that it is known from the course of general algebra.

PROOF. Let's begin the proof of the theorem 5.2 by writing the relationships (5.5) and (5.7) in a brief symbolic form:

$$\tilde{\mathbf{e}}_i = \sum_{k=1}^n S_i^k \cdot \mathbf{e}_k, \quad \mathbf{e}_j = \sum_{i=1}^n T_j^i \cdot \tilde{\mathbf{e}}_i. \quad (5.8)$$

Then we substitute the first relationship (5.8) into the second one. This yields:

$$\mathbf{e}_j = \sum_{i=1}^n T_j^i \cdot \left(\sum_{k=1}^n S_i^k \cdot \mathbf{e}_k \right) = \sum_{k=1}^n \left(\sum_{i=1}^n S_i^k T_j^i \right) \cdot \mathbf{e}_k. \quad (5.9)$$

The symbol δ_j^k , which is called the *Kronecker symbol*, is used for denoting the following numeric array:

$$\delta_j^k = \begin{cases} 1 & \text{for } k = j, \\ 0 & \text{for } k \neq j. \end{cases} \quad (5.10)$$

We apply the Kronecker symbol determined in (5.10) in order to transform left hand side of the equality (5.9):

$$\mathbf{e}_j = \sum_{k=1}^n \delta_j^k \cdot \mathbf{e}_k. \quad (5.11)$$

Both equalities (5.11) and (5.9) represent the same vector $\mathbf{e}_j$ expanded in the same basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. Due to the theorem 5.1 on the uniqueness of the expansion of a vector in a basis we have the equality

$$\sum_{i=1}^n S_i^k T_j^i = \delta_j^k.$$

It is easy to note that this equality is equivalent to the matrix equality $ST = 1$. The theorem is proved. $\square$

COROLLARY. *The direct transition matrix S and the inverse transition matrix T both are non-degenerate matrices and $\det S \det T = 1$.*

PROOF. The relationship $\det S \det T = 1$ follows from the matrix equality $ST = 1$, which was proved just above. This fact is well known from the course of general algebra. If the product of two numbers is equal to unity, then none of these two numbers can be equal to zero:

$$\det S \neq 0, \quad \det T \neq 0.$$

This proves the non-degeneracy of transition matrices S and T . The corollary is proved. $\square$

THEOREM 5.3. *Every non-degenerate $n \times n$ matrix S can be obtained as a transition matrix for passing from some basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ to some other basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ in a linear vector space V of the dimension n .*

PROOF. Let's choose an arbitrary $\mathbf{e}_1, \dots, \mathbf{e}_n$ basis in V and fix it. Then let's determine the other n vectors $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ by means of the relationships (5.5) and prove that they are linearly independent. For this purpose we consider a linear combination of these vectors that is equal to zero:

$$\alpha^1 \cdot \tilde{\mathbf{e}}_1 + \dots + \alpha^n \cdot \tilde{\mathbf{e}}_n = \mathbf{0}. \quad (5.12)$$

Substituting (5.5) into this equality, one can transform it to the following one:

$$\left(\sum_{i=1}^n S_i^1 \alpha^i \right) \cdot \mathbf{e}_1 + \dots + \left(\sum_{i=1}^n S_i^n \alpha^i \right) \cdot \mathbf{e}_n = 0.$$

Since the basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ are linearly independent, it follows that all sums enclosed within the brackets in the above equality are equal to zero. Writing

This is exactly the first transformation formula (5.14). The second formula (5.14) is proved similarly. $\square$

§ 6. Intersections and sums of subspaces.

Suppose that we have a certain number of subspaces in a linear vector space V . In order to designate this fact we write $U_i \subset V$, where $i \in I$. The number of subspaces can be finite or infinite enumerable, then they can be enumerated by the positive integers. However, in general case we should enumerate the subspaces by the elements of some indexing set I , which can be finite, infinite enumerable, or even non-enumerable. Let's denote by U and by S the intersection and the union of all subspaces that we consider:

$$U = \bigcap_{i \in I} U_i, \quad S = \bigcup_{i \in I} U_i. \quad (6.1)$$

THEOREM 6.1. *The intersection of an arbitrary number of subspaces in a linear vector space V is a subspace in V .*

PROOF. The set U in (6.1) is not empty since zero vector is an element of each subspace U_i . Let's verify the conditions (1) and (2) from the definition 2.2 for U .

Suppose that $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}$ are the vectors from the subset U . Then they belong to U_i for each $i \in I$. However, U_i is a subspace, hence, $u_1 + u_2 \in U_i$ and $\alpha \cdot u \in U_i$ for any $i \in I$ and for any $\alpha \in \mathbb{K}$. Therefore, $u_1 + u_2 \in U$ and $\alpha \cdot u \in U$. The theorem is proved. $\square$

In general, the subset S in (6.1) is not a subspace. Therefore we need to introduce the following concept.

DEFINITION 6.1. The linear span of the union of subspaces U_i , $i \in I$, is called the *sum* of these subspaces.

For to denote the sum of subspaces $W = \langle S \rangle$ we use the standard summation sign:

$$W = \left\langle \bigcup_{i \in I} U_i \right\rangle = \sum_{i \in I} U_i.$$

THEOREM 6.2. *A vector $\mathbf{w}$ of a linear vector space V belongs to the sum of subspaces U_i , $i \in I$, if and only if it is represented as a sum of finite number of vectors each of which is taken from some subspace U_i :*

$$\mathbf{w} = \mathbf{u}_{i_1} + \dots + \mathbf{u}_{i_k}, \text{ where } \mathbf{u}_i \in U_i. \quad (6.2)$$

PROOF. Let S be the union of subspaces $U_i \subset V$, $i \in I$. Suppose that $\mathbf{w} \in W$. Then $\mathbf{w}$ is a linear combination of finite number of vectors taken from S :

$$\mathbf{w} = \alpha_1 \cdot \mathbf{s}_1 + \dots + \alpha_k \cdot \mathbf{s}_k.$$

But S is the union of subspaces U_i . Therefore, $\mathbf{s}_m \in U_{i_m}$ and $\alpha_m \cdot \mathbf{s}_m = \mathbf{u}_{i_m} \in U_{i_m}$, where $m = 1, \dots, k$. This leads to the equality (6.2) for the vector $\mathbf{w}$.

Conversely, suppose that $\mathbf{w}$ is a vector given by formula (6.2). Then $u_{i_m} \in U_{i_m}$ and $U_{i_m} \subset S$, i.e. $u_{i_m} \in S$. Therefore, the vector $\mathbf{w}$ belongs to the linear span of S . The theorem is proved. $\square$

DEFINITION 6.2. The sum W of subspaces U_i , $i \in I$, is called the *direct sum*, if for any vector $\mathbf{w} \in W$ the expansion (6.2) is unique. In this case for the direct sum of subspaces we use the special notation:

$$W = \bigoplus_{i \in I} U_i.$$

THEOREM 6.3. Let $W = U_1 + \dots + U_k$ be the sum a finite number of finite-dimensional subspaces. The dimension of W is equal to the sum of dimensions of the subspaces U_i if and only if W is the direct sum: $W = U_1 \oplus \dots \oplus U_k$.

PROOF. Let's choose a basis in each subspace U_i . Suppose that $\dim U_i = s_i$ and let $\mathbf{e}_{i1}, \dots, \mathbf{e}_{is_i}$ be a basis in U_i . Let's join the vectors of all bases into one system ordering them alphabetically:

$$\mathbf{e}_{11}, \dots, \mathbf{e}_{1s_1}, \dots, \mathbf{e}_{k1}, \dots, \mathbf{e}_{ks_k}. \quad (6.3)$$

Due to the equality $W = U_1 + \dots + U_k$ for an arbitrary vector $\mathbf{w}$ of the subspace W we have the expansion (6.2):

$$\mathbf{w} = \mathbf{u}_1 + \dots + \mathbf{u}_k, \text{ where } \mathbf{u}_i \in U_i. \quad (6.4)$$

Expanding each vector $\mathbf{u}_i$ of (6.4) in the basis of corresponding subspace U_i , we get the expansion of $\mathbf{w}$ in vectors of the system (6.3). Hence, (6.3) is a spanning system of vectors in W (though, in general case it is not a minimal spanning system).

If $\dim W = \dim U_1 + \dots + \dim U_k$, then the number of vectors in (6.3) cannot be reduced. Therefore (6.3) is a basis in W . From any expansion (6.4) we can derive the following expansion of the vector $\mathbf{w}$ in the basis (6.3):

$$\mathbf{w} = \left(\sum_{j=1}^{s_1} \alpha_{1j} \cdot \mathbf{e}_{1j} \right) + \dots + \left(\sum_{j=1}^{s_k} \alpha_{kj} \cdot \mathbf{e}_{kj} \right). \quad (6.5)$$

The sums enclosed into the round brackets in (6.5) are determined by the expansions of the vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ in the bases of corresponding subspaces $U_1, \dots, U_k$:

$$u_i = \sum_{j=1}^{s_i} \alpha_{ij} \cdot \mathbf{e}_{ij}. \quad (6.6)$$

Due to (6.6) the existence of two different expansions (6.4) for some vector $\mathbf{w}$ would mean the existence of two different expansions (6.5) of this vector in the basis (6.3). Hence, the expansion (6.4) is unique and the sum of subspaces $W = U_1 + \dots + U_k$ is the direct sum.

Conversely, suppose that $W = U_1 \oplus \dots \oplus U_k$. We know that the vectors (6.3) span the subspace W . Let's prove that they are linearly independent. For this purpose we consider a linear combination of these vectors being equal to zero:

$$\mathbf{0} = \left(\sum_{j=1}^{s_1} \alpha_{1j} \cdot \mathbf{e}_{1j} \right) + \dots + \left(\sum_{j=1}^{s_k} \alpha_{kj} \cdot \mathbf{e}_{kj} \right). \quad (6.7)$$

Let's denote by $\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_k$ the values of sums enclosed into the round brackets in (6.7). It is easy to see that $\tilde{\mathbf{u}}_i \in U_i$, therefore, (6.7) is an expansion of the form (6.4) for the vector $\mathbf{w} = \mathbf{0}$. But $\mathbf{0} = \mathbf{0} + \dots + \mathbf{0}$ and $\mathbf{0} \in U_i$. This is another expansion for the vector $\mathbf{w} = \mathbf{0}$. However, $W = U_1 \oplus \dots \oplus U_k$, therefore, the expansion $\mathbf{0} = \mathbf{0} + \dots + \mathbf{0}$ is unique expansion of the form (6.4) for zero vector $\mathbf{w} = \mathbf{0}$. Then we have the equalities

$$\mathbf{0} = \sum_{j=1}^{s_i} \alpha_{ij} \cdot \mathbf{e}_{ij} \text{ for all } i = 1, \dots, k.$$

It's clear that these equalities are the expansions of zero vector in the bases of the subspaces U_i . Hence, $\alpha_{ij} = 0$. This means that the linear combination (6.7) is trivial, and (6.3) is a linearly independent system of vectors. Thus, being a spanning system and being linearly independent, the system of vectors (6.3) is a basis of W . Now we can find the dimension of the subspace W by counting the number of vectors in (6.3): $\dim W = s_1 + \dots + s_k = \dim U_1 + \dots + \dim U_k$. The theorem is proved. $\square$

Note. If the sum of subspaces $W = U_1 + \dots + U_k$ is not necessarily the direct sum, the vectors (6.3), nevertheless, form a spanning system in W . But they do not necessarily form a linearly independent system in this case. Therefore, we have

$$\dim W \leq \dim U_1 + \dots + \dim U_k. \quad (6.8)$$

Sharpening this inequality in general case is sufficiently complicated. We shall do it for the case of two subspaces.

THEOREM 6.4. *The dimension of the sum of two arbitrary finite-dimensional subspaces U_1 and U_2 in a linear vector space V is equal to the sum of their dimensions minus the dimension of their intersection:*

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2). \quad (6.9)$$

PROOF. From the inclusion $U_1 \cap U_2 \subset U_1$ and from the inequality (6.8) we conclude that all subspaces considered in the theorem are finite-dimensional. Let's denote $\dim(U_1 \cap U_2) = s$ and choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ in the intersection $U_1 \cap U_2$.

Due to the inclusion $U_1 \cap U_2 \subset U_1$ we can apply the theorem 4.8 on completing the basis. This theorem says that we can complete the basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ of the intersection $U_1 \cap U_2$ up to a basis $\mathbf{e}_1, \dots, \mathbf{e}_s, \mathbf{e}_{s+1}, \dots, \mathbf{e}_{s+p}$ in U_1 . For the dimension of U_1 , we have $\dim U_1 = s + p$. In a similar way, due to the inclusion $U_1 \cap U_2 \subset U_2$ we can construct a basis $\mathbf{e}_1, \dots, \mathbf{e}_s, \mathbf{e}_{s+p+1}, \dots, \mathbf{e}_{s+p+q}$ in U_2 . For the dimension of U_2 this yields $\dim U_2 = s + q$.

Now let's join together the two bases constructed above with the use of theorem 4.8 and consider the total set of vectors in them:

$$\mathbf{e}_1, \dots, \mathbf{e}_s, \mathbf{e}_{s+1}, \dots, \mathbf{e}_{s+p}, \mathbf{e}_{s+p+1}, \dots, \mathbf{e}_{s+p+q}. \quad (6.10)$$

Let's prove that these vectors (6.10) form a basis in the sum of subspaces $U_1 + U_2$. Let $\mathbf{w}$ be some arbitrary vector in $U_1 + U_2$. The relationship (6.2) for this vector is written as $\mathbf{w} = \mathbf{u}_1 + \mathbf{u}_2$. Let's expand the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ in the above two bases of the subspaces U_1 and U_2 respectively:

$$\begin{aligned} \mathbf{u}_1 &= \sum_{i=1}^s \alpha_i \cdot \mathbf{e}_i + \sum_{j=1}^p \beta_{s+j} \cdot \mathbf{e}_{s+j}, \\ \mathbf{u}_2 &= \sum_{i=1}^s \tilde{\alpha}_i \cdot \mathbf{e}_i + \sum_{j=1}^q \gamma_{s+p+j} \cdot \mathbf{e}_{s+p+j}. \end{aligned}$$

Adding these two equalities, we find that the vector $\mathbf{w}$ is linearly expressed through the vectors (6.10). Hence, (6.10) is a spanning system of vectors in $U_1 + U_2$.

In order to prove that (6.10) is a linearly independent system of vectors we consider a linear combination of these vectors being equal to zero:

$$\sum_{i=1}^{s+p} \alpha_i \cdot \mathbf{e}_i + \sum_{i=1}^q \alpha_{s+p+i} \cdot \mathbf{e}_{s+p+i} = \mathbf{0}. \quad (6.11)$$

Then we transform this equality by moving the second sum to the right hand side:

$$\sum_{i=1}^{s+p} \alpha_i \cdot \mathbf{e}_i = - \sum_{i=1}^q \alpha_{s+p+i} \cdot \mathbf{e}_{s+p+i}.$$

Let's denote by $\mathbf{u}$ the value of left and right sides of this equality. Then for the vector $\mathbf{u}$ we get the following two expressions:

$$\mathbf{u} = \sum_{i=1}^{s+p} \alpha_i \cdot \mathbf{e}_i, \quad \mathbf{u} = - \sum_{i=1}^q \alpha_{s+p+i} \cdot \mathbf{e}_{s+p+i}. \quad (6.12)$$

Because of the first expression (6.12) we have $\mathbf{u} \in U_1$, while the second expression (6.12) yields $\mathbf{u} \in U_2$. Hence, $\mathbf{u} \in U_1 \cap U_2$. This means that we can expand the vector $\mathbf{u}$ in the basis $\mathbf{e}_1, \dots, \mathbf{e}_s$:

$$\mathbf{u} = \sum_{i=1}^s \beta_i \cdot \mathbf{e}_i. \quad (6.13)$$

Comparing this expansion with the second expression (6.12), we find that

$$\sum_{i=1}^s \beta_i \cdot \mathbf{e}_i + \sum_{i=1}^q \alpha_{s+p+i} \cdot \mathbf{e}_{s+p+i} = \mathbf{0}. \quad (6.14)$$

Note that the vectors $\mathbf{e}_1, \dots, \mathbf{e}_s, \mathbf{e}_{s+p+1}, \dots, \mathbf{e}_{s+p+q}$ form a basis in U_2 . They are linearly independent. Therefore, all coefficients in (6.14) are equal to zero. In particular, we have the following equalities:

$$\alpha_{s+p+1} = \dots = \alpha_{s+p+q} = 0. \quad (6.15)$$

Moreover, $\beta_1 = \dots = \beta_s = 0$. Due to (6.13) this means that $\mathbf{u} = 0$. Now from the first expansion (6.12) we get the equality

$$\sum_{i=1}^{s+p} \alpha_i \cdot \mathbf{e}_i = 0.$$

Since $\mathbf{e}_1, \dots, \mathbf{e}_s, \mathbf{e}_{s+1}, \dots, \mathbf{e}_{s+p}$ are linearly independent vectors, all coefficients α_i in the above equality should be zero:

$$\alpha_1 = \dots = \alpha_s = \alpha_{s+1} = \dots = \alpha_{s+p} = 0. \quad (6.16)$$

Combining (6.15) and (6.16), we see that the linear combination (6.11) is trivial. This means that the vectors (6.10) are linearly independent. Hence, they form a basis in $U_1 + U_2$. For the dimension of the subspace $U_1 + U_2$ this yields

$$\begin{aligned} \dim(U_1 + U_2) &= s + p + q = (s + p) + (s + q) - s = \\ &= \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2). \end{aligned}$$

Thus, the relationship (6.9) and the theorem 6.4 in whole is proved. $\square$

§ 7. Cosets of a subspace. The concept of factorspace.

Let V be a linear vector space and let U be a subspace in it. A *coset* of the subspace U determined by a vector $\mathbf{v} \in V$ is the following set of vectors¹:

$$\text{Cl}_U(\mathbf{v}) = \{\mathbf{w} \in V : \mathbf{w} - \mathbf{v} \in U\}. \quad (7.1)$$

The vector $\mathbf{v}$ in (7.1) is called a *representative* of the coset (7.1). The coset $\text{Cl}_U(\mathbf{v})$ is a very simple thing, it is obtained by adding the vector $\mathbf{v}$ with all vectors of the subspace U . The coset represented by zero vector is the especially simple thing since $\text{Cl}_U(\mathbf{0}) = U$. It is called a *zero coset*.

THEOREM 7.1. *The cosets of a subspace U in a linear vector space V possess the following properties:*

- (1) $\mathbf{a} \in \text{Cl}_U(\mathbf{a})$ for any $\mathbf{a} \in V$;
- (2) if $\mathbf{a} \in \text{Cl}_U(\mathbf{b})$, then $\mathbf{b} \in \text{Cl}_U(\mathbf{a})$;
- (3) if $\mathbf{a} \in \text{Cl}_U(\mathbf{b})$ and $\mathbf{b} \in \text{Cl}_U(\mathbf{c})$, then $\mathbf{a} \in \text{Cl}_U(\mathbf{c})$.

PROOF. The first proposition is obvious. Indeed, the difference $\mathbf{a} - \mathbf{a}$ is equal to zero vector, which is an element of any subspace: $\mathbf{a} - \mathbf{a} = \mathbf{0} \in U$. Hence, due to the formula (7.1), which is the formal definition of cosets, we have $\mathbf{a} \in \text{Cl}_U(\mathbf{a})$.

¹ We used the sign Cl for cosets since in Russia they are called *adjacency classes*.

Let $\mathbf{a} \in \text{Cl}_U(\mathbf{b})$. Then $\mathbf{a} - \mathbf{b} \in U$. For $\mathbf{b} - \mathbf{a}$, we have $\mathbf{b} - \mathbf{a} = (-1) \cdot (\mathbf{a} - \mathbf{b})$. Therefore, $\mathbf{b} - \mathbf{a} \in U$ and $\mathbf{b} \in \text{Cl}_U(\mathbf{a})$ (see formula (7.1) and the definition 2.2). The second proposition is proved.

Let $\mathbf{a} \in \text{Cl}_U(\mathbf{b})$ and $\mathbf{b} \in \text{Cl}_U(\mathbf{c})$. Then $\mathbf{a} - \mathbf{b} \in U$ and $\mathbf{b} - \mathbf{c} \in U$. Note that $\mathbf{a} - \mathbf{c} = (\mathbf{a} - \mathbf{b}) + (\mathbf{b} - \mathbf{c})$. Hence, $\mathbf{a} - \mathbf{c} \in U$ and $\mathbf{a} \in \text{Cl}_U(\mathbf{c})$ (see formula (7.1) and the definition 2.2 again). The third proposition is proved. This completes the proof of the theorem in whole. $\square$

Let $\mathbf{a} \in \text{Cl}_U(\mathbf{b})$. This condition establishes some kind of dependence between two vectors $\mathbf{a}$ and $\mathbf{b}$. This dependence is not strict: the condition $\mathbf{a} \in \text{Cl}_U(\mathbf{b})$ does not exclude the possibility that $\mathbf{a}' \in \text{Cl}_U(\mathbf{b})$ for some other vector $\mathbf{a}'$. Such non-strict dependences in mathematics are described by the concept of *binary relation* (see details in [1] and [4]). Let's write $\mathbf{a} \sim \mathbf{b}$ as an abbreviation for $\mathbf{a} \in \text{Cl}_U(\mathbf{b})$. Then the theorem 7.1 reveals the following properties of the binary relation $\mathbf{a} \sim \mathbf{b}$, which is introduced just above:

- (1) *reflexivity*: $\mathbf{a} \sim \mathbf{a}$;
- (2) *symmetry*: $\mathbf{a} \sim \mathbf{b}$ implies $\mathbf{b} \sim \mathbf{a}$;
- (3) *transitivity*: $\mathbf{a} \sim \mathbf{b}$ and $\mathbf{b} \sim \mathbf{c}$ implies $\mathbf{a} \sim \mathbf{c}$.

A binary relation possessing the properties of reflexivity, symmetry, and transitivity is called an *equivalence relation*. Each equivalence relation determined in a set V partitions this set into a union of mutually non-intersecting subsets, which are called the *equivalence classes*:

$$\text{Cl}(\mathbf{v}) = \{\mathbf{w} \in V : \mathbf{w} \sim \mathbf{v}\}. \quad (7.2)$$

In our particular case the formal definition (7.2) coincides with the formal definition (7.1). In order to keep the completeness of presentation we shall not use the notation $\mathbf{a} \sim \mathbf{b}$ in place of $\mathbf{a} \in \text{Cl}_U(\mathbf{b})$ anymore, and we shall not refer to the theory of binary relations (though it is simple and well-known). Instead of this we shall derive the result on partitioning V into the mutually non-intersecting cosets from the following theorem.

THEOREM 7.2. *If two cosets $\text{Cl}_U(\mathbf{a})$ and $\text{Cl}_U(\mathbf{b})$ of a subspace $U \subset V$ are intersecting, then they do coincide.*

PROOF. Assume that the intersection of two cosets $\text{Cl}_U(\mathbf{a})$ and $\text{Cl}_U(\mathbf{b})$ is not empty. Then there is an element $\mathbf{c}$ belonging to both of them: $\mathbf{c} \in \text{Cl}_U(\mathbf{a})$ and $\mathbf{c} \in \text{Cl}_U(\mathbf{b})$. Due to the proposition (2) of the above theorem 7.1 we derive $\mathbf{b} \in \text{Cl}_U(\mathbf{c})$. Combining $\mathbf{b} \in \text{Cl}_U(\mathbf{c})$ and $\mathbf{c} \in \text{Cl}_U(\mathbf{a})$ and applying the proposition (3) of the theorem 7.1, we get $\mathbf{b} \in \text{Cl}_U(\mathbf{a})$. The opposite inclusion $\mathbf{a} \in \text{Cl}_U(\mathbf{b})$ then is obtained by applying the proposition (2) of the theorem 7.1.

Let's prove that two cosets $\text{Cl}_U(\mathbf{a})$ and $\text{Cl}_U(\mathbf{b})$ do coincide. For this purpose let's consider an arbitrary vector $\mathbf{x} \in \text{Cl}_U(\mathbf{a})$. From $\mathbf{x} \in \text{Cl}_U(\mathbf{a})$ and $\mathbf{a} \in \text{Cl}_U(\mathbf{b})$ we derive $\mathbf{x} \in \text{Cl}_U(\mathbf{b})$. Hence, $\text{Cl}_U(\mathbf{a}) \subset \text{Cl}_U(\mathbf{b})$. The opposite inclusion $\text{Cl}_U(\mathbf{b}) \subset \text{Cl}_U(\mathbf{a})$ is proved similarly. From these two inclusions we derive $\text{Cl}_U(\mathbf{a}) = \text{Cl}_U(\mathbf{b})$. The theorem is proved. $\square$

The set of all cosets of a subspace U in a linear vector space V is called the *factorset* or *quotient set* V/U . Due to the theorem proved just above any two

different cosets Q_1 and Q_2 from the factorset V/U have the empty intersection $Q_1 \cap Q_2 = \emptyset$, while the union of all cosets coincides with V :

$$V = \bigcup_{Q \in V/U} Q.$$

This equality is a consequence of the fact that any vector $\mathbf{v} \in V$ is an element of some coset: $\mathbf{v} \in Q$. This coset Q is determined by $\mathbf{v}$ according to the formula $Q = \text{Cl}_U(\mathbf{v})$. For this reason the following theorem is a simple reformulation of the definition of cosets.

THEOREM 7.3. *Two vectors $\mathbf{v}$ and $\mathbf{w}$ belong to the same coset of a subspace U if and only if their difference $\mathbf{v} - \mathbf{w}$ is a vector of U .*

DEFINITION 7.1. Let Q_1 and Q_2 be two cosets of a subspace U . The *sum* of cosets Q_1 and Q_2 is a coset Q of the subspace U determined by the equality $Q = \text{Cl}_U(\mathbf{v}_1 + \mathbf{v}_2)$, where $\mathbf{v}_1 \in Q_1$ and $\mathbf{v}_2 \in Q_2$.

DEFINITION 7.2. Let Q be a coset of a subspace U . The product of Q and a number $\alpha \in \mathbb{K}$ is a coset P of the subspace U determined by the relationship $P = \text{Cl}_U(\alpha \cdot \mathbf{v})$, where $\mathbf{v} \in Q$.

For the addition of cosets and for the multiplication of them by numbers we use the same signs of algebraic operations as in case of vectors, i. e. $Q = Q_1 + Q_2$ and $P = \alpha \cdot Q$. The definitions 7.1 and 7.2 can be expressed by formulas

$$\begin{aligned} \text{Cl}_U(\mathbf{v}_1) + \text{Cl}_U(\mathbf{v}_2) &= \text{Cl}_U(\mathbf{v}_1 + \mathbf{v}_2), \\ \alpha \cdot \text{Cl}_U(\mathbf{v}) &= \text{Cl}_U(\alpha \cdot \mathbf{v}). \end{aligned} \tag{7.3}$$

These definitions require some comments. Indeed, the coset $Q = Q_1 + Q_2$ in the definition 7.1 and the coset $P = \alpha \cdot Q$ in the definition 7.2 both are determined using some representative vectors $\mathbf{v}_1 \in Q_1$, $\mathbf{v}_2 \in Q_2$, and $\mathbf{v} \in Q$. The choice of a representative vector in a coset is not unique; therefore, we need especially to prove the uniqueness of the results of algebraic operations determined in the definitions 7.1 and 7.2. This proof is called the *proof of correctness*.

THEOREM 7.4. *The definitions 7.1 and 7.2 are correct and the results of the algebraic operations of coset addition and of coset multiplication by numbers do not depend on the choice of representatives in cosets.*

PROOF. For the beginning we study the operation of coset addition. Let's take consider two different choices of representatives within cosets Q_1 and Q_2 . Let $\mathbf{v}_1, \tilde{\mathbf{v}}_1$ be two vectors of Q_1 and let $\mathbf{v}_2, \tilde{\mathbf{v}}_2$ be two vectors of Q_2 . Then due to the theorem 7.3 we have the following two equalities:

$$\tilde{\mathbf{v}}_1 - \mathbf{v}_1 \in U, \quad \tilde{\mathbf{v}}_2 - \mathbf{v}_2 \in U.$$

Hence, $(\tilde{\mathbf{v}}_1 + \tilde{\mathbf{v}}_2) - (\mathbf{v}_1 + \mathbf{v}_2) = (\tilde{\mathbf{v}}_1 - \mathbf{v}_1) + (\tilde{\mathbf{v}}_2 - \mathbf{v}_2) \in U$. This means that the cosets determined by vectors $\tilde{\mathbf{v}}_1 + \tilde{\mathbf{v}}_2$ and $\mathbf{v}_1 + \mathbf{v}_2$ do coincide with each other:

$$\text{Cl}_U(\tilde{v}_1 + \tilde{v}_2) = \text{Cl}_U(v_1 + v_2).$$

This proves the correctness of the definition 7.1 for the operation of coset addition.

Now let's consider two different representatives $\mathbf{v}$ and $\tilde{\mathbf{v}}$ within the coset Q . Then $\tilde{\mathbf{v}} - \mathbf{v} \in U$. Hence, $\alpha \cdot \tilde{\mathbf{v}} - \alpha \cdot \mathbf{v} = \alpha \cdot (\tilde{\mathbf{v}} - \mathbf{v}) \in U$. This yields

$$\text{Cl}_U(\alpha \cdot \tilde{\mathbf{v}}) = \text{Cl}_U(\alpha \cdot \mathbf{v}),$$

which proves the correctness of the definition 7.2 for the operation of multiplication of cosets by numbers. $\square$

THEOREM 7.5. *The factorset V/U of a linear vector space V over a subspace U equipped with algebraic operations (7.3) is a linear vector space. This space is called the factorspace or the quotient space of the space V over its subspace U .*

PROOF. The proof of this theorem consists in verifying the axioms (1)-(8) of a linear vector space for V/U . The commutativity and associativity axioms for the operation of coset addition follow from the following calculations:

$$\begin{aligned} \text{Cl}_U(\mathbf{v}_1) + \text{Cl}_U(\mathbf{v}_2) &= \text{Cl}_U(\mathbf{v}_1 + \mathbf{v}_2) = \\ &= \text{Cl}_U(\mathbf{v}_2 + \mathbf{v}_1) = \text{Cl}_U(\mathbf{v}_2) + \text{Cl}_U(\mathbf{v}_1), \\ (\text{Cl}_U(\mathbf{v}_1) + \text{Cl}_U(\mathbf{v}_2)) + \text{Cl}_U(\mathbf{v}_3) &= \text{Cl}_U(\mathbf{v}_1 + \mathbf{v}_2) + \text{Cl}_U(\mathbf{v}_3) = \\ &= \text{Cl}_U((\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3) = \text{Cl}_U(\mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)) = \\ &= \text{Cl}_U(\mathbf{v}_1) + \text{Cl}_U(\mathbf{v}_2 + \mathbf{v}_3) = \text{Cl}_U(\mathbf{v}_1) + (\text{Cl}_U(\mathbf{v}_2) + \text{Cl}_U(\mathbf{v}_3)). \end{aligned}$$

In essential, they follow from the corresponding axioms for the operation of vector addition (see definition 2.1).

In order to verify the axiom (3) we should have a zero element in V/U . The zero coset $\mathbf{0} = \text{Cl}_U(\mathbf{0})$ is the best pretender for this role:

$$\text{Cl}_U(\mathbf{v}) + \text{Cl}_U(\mathbf{0}) = \text{Cl}_U(\mathbf{v} + \mathbf{0}) = \text{Cl}_U(\mathbf{v}).$$

In verifying the axiom (4) we should indicate the opposite coset Q' for a coset $Q = \text{Cl}_U(\mathbf{v})$. We define it as follows: $Q' = \text{Cl}_U(\mathbf{v}')$. Then

$$Q + Q' = \text{Cl}_U(\mathbf{v}) + \text{Cl}_U(\mathbf{v}') = \text{Cl}_U(\mathbf{v} + \mathbf{v}') = \text{Cl}_U(\mathbf{0}) = \mathbf{0}.$$

The rest axioms (5)-(8) are verified by direct calculations on the base of formula (7.3) for coset operations. Here are these calculations:

$$\begin{aligned} \alpha \cdot (\text{Cl}_U(\mathbf{v}_1) + \text{Cl}_U(\mathbf{v}_2)) &= \alpha \cdot \text{Cl}_U(\mathbf{v}_1 + \mathbf{v}_2) = \\ &= \text{Cl}_U(\alpha \cdot (\mathbf{v}_1 + \mathbf{v}_2)) = \text{Cl}_U(\alpha \cdot \mathbf{v}_1 + \alpha \cdot \mathbf{v}_2) = \\ &= \text{Cl}_U(\alpha \cdot \mathbf{v}_1) + \text{Cl}_U(\alpha \cdot \mathbf{v}_2) = \alpha \cdot \text{Cl}_U(\mathbf{v}_1) + \alpha \cdot \text{Cl}_U(\mathbf{v}_2), \\ (\alpha + \beta) \cdot \text{Cl}_U(\mathbf{v}) &= \text{Cl}_U((\alpha + \beta) \cdot \mathbf{v}) = \text{Cl}_U(\alpha \cdot \mathbf{v} + \beta \cdot \mathbf{v}) = \\ &= \text{Cl}_U(\alpha \cdot \mathbf{v}) + \text{Cl}_U(\beta \cdot \mathbf{v}) = \alpha \cdot \text{Cl}_U(\mathbf{v}) + \beta \cdot \text{Cl}_U(\mathbf{v}), \\ \alpha \cdot (\beta \cdot \text{Cl}_U(\mathbf{v})) &= \alpha \cdot \text{Cl}_U(\beta \cdot \mathbf{v}) = \text{Cl}_U(\alpha \cdot (\beta \cdot \mathbf{v})) = \\ &= \text{Cl}_U((\alpha\beta) \cdot \mathbf{v}) = (\alpha\beta) \cdot \text{Cl}_U(\mathbf{v}), \\ 1 \cdot \text{Cl}_U(\mathbf{v}) &= \text{Cl}_U(1 \cdot \mathbf{v}) = \text{Cl}_U(\mathbf{v}). \end{aligned}$$

The above equalities complete the verification of the fact that the factorset V/U possesses the structure of a linear vector space. $\square$

Note that verifying the axiom (4) we have defined the opposite coset Q' for a coset $Q = Cl_U(\mathbf{v})$ by means of the relationship $Q' = Cl_U(\mathbf{v}')$, where $\mathbf{v}'$ is the opposite vector for $\mathbf{v}$. One could check the correctness of this definition. However, this is not necessary since due to the property (10), see theorem 2.1, the opposite coset Q' for Q is unique.

The concept of factorspace is equally applicable to finite-dimensional and to infinite-dimensional spaces V . The finite or infinite dimensionality of a subspace U also makes no difference. The only simplification in finite-dimensional case is that we can calculate the dimension of the factorspace V/U .

THEOREM 7.6. *If a linear vector space V is finite-dimensional, then for any its subspace U the factorspace V/U also is finite-dimensional and its dimension is determined by the following formula:*

$$\dim U + \dim(V/U) = \dim V. \quad (7.4)$$

PROOF. If $U = V$ then the factorspace V/U consists of zero coset only: $V/U = \{\mathbf{0}\}$. The dimension of such zero space is equal to zero. Hence, the equality (7.4) in this trivial case is fulfilled.

Let's consider a nontrivial case $U \subsetneq V$. Due to the theorem 4.5 the subspace U is finite-dimensional. Denote $\dim V = n$ and $\dim U = s$, then $s < n$. Let's choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ in U and, according to the theorem 4.8, complete it with vectors $\mathbf{e}_{s+1}, \dots, \mathbf{e}_n$ up to a basis in V . For each of complementary vectors $\mathbf{e}_{s+1}, \dots, \mathbf{e}_n$ we consider the corresponding coset of a subspace U :

$$\mathbf{E}_1 = Cl_U(\mathbf{e}_{s+1}), \dots, \mathbf{E}_{n-s} = Cl_U(\mathbf{e}_n). \quad (7.5)$$

Now let's show that the cosets (7.5) span the factorspace V/U . Indeed, let Q be an arbitrary coset in V/U and let $\mathbf{v} \in Q$ be some representative vector of this coset. Let's expand the vector $\mathbf{v}$ in the above basis of V :

$$\mathbf{v} = (\alpha_1 \cdot \mathbf{e}_1 + \dots + \alpha_s \cdot \mathbf{e}_s) + \beta_1 \cdot \mathbf{e}_{s+1} + \dots + \beta_{n-s} \cdot \mathbf{e}_n.$$

Let's denote by $\mathbf{u}$ the initial part of this expansion: $\mathbf{u} = \alpha_1 \cdot \mathbf{e}_1 + \dots + \alpha_s \cdot \mathbf{e}_s$. It is clear that $\mathbf{u} \in U$. Then we can write

$$\mathbf{v} = \mathbf{u} + \beta_1 \cdot \mathbf{e}_{s+1} + \dots + \beta_{n-s} \cdot \mathbf{e}_n.$$

Since $\mathbf{u} \in U$, we have $Cl_U(\mathbf{u}) = \mathbf{0}$. For the coset $Q = Cl_U(\mathbf{v})$ this equality yields $Q = \beta_1 \cdot Cl_U(\mathbf{e}_{s+1}) + \dots + \beta_{n-s} \cdot Cl_U(\mathbf{e}_n)$. Hence, we have

$$Q = \beta_1 \cdot \mathbf{E}_1 + \dots + \beta_{n-s} \cdot \mathbf{E}_{n-s}.$$

This means that $\mathbf{E}_1, \dots, \mathbf{E}_{n-s}$ is a finite spanning system in V/U . Therefore, V/U is a finite-dimensional linear vector space. To determine its dimension we

shall prove that the cosets (7.5) are linearly independent. Indeed, let's consider a linear combination of these cosets being equal to zero:

$$\gamma_1 \cdot \mathbf{E}_1 + \dots + \gamma_{n-s} \cdot \mathbf{E}_{n-s} = \mathbf{0}. \quad (7.6)$$

Passing from cosets to their representative vectors, from (7.6) we derive

$$\begin{aligned} \gamma_1 \cdot \text{Cl}_U(\mathbf{e}_{s+1}) + \dots + \gamma_{n-s} \cdot \text{Cl}_U(\mathbf{e}_n) &= \\ = \text{Cl}_U(\gamma_1 \cdot \mathbf{e}_{s+1} + \dots + \gamma_{n-s} \cdot \mathbf{e}_n) &= \text{Cl}_U(\mathbf{0}). \end{aligned}$$

Let's denote $\mathbf{u} = \gamma_1 \cdot \mathbf{e}_{s+1} + \dots + \gamma_{n-s} \cdot \mathbf{e}_n$. From the above equality for this vector we get $\text{Cl}_U(\mathbf{u}) = \text{Cl}_U(\mathbf{0})$, which means $\mathbf{u} \in U$. Let's expand $\mathbf{u}$ in the basis of subspace U : $\mathbf{u} = \alpha_1 \cdot \mathbf{e}_1 + \dots + \alpha_s \cdot \mathbf{e}_s$. Then, equating two expressions for the vector $\mathbf{u}$, we get the following equality:

$$-\alpha_1 \cdot \mathbf{e}_1 - \dots - \alpha_s \cdot \mathbf{e}_s + \gamma_1 \cdot \mathbf{e}_{s+1} + \dots + \gamma_{n-s} \cdot \mathbf{e}_n = \mathbf{0}.$$

This is the linear combination of basis vectors of V , which is equal to zero. Basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ are linearly independent. Hence, this linear combination is trivial and $\gamma_1 = \dots = \gamma_{n-s} = 0$. This proves the triviality of linear combination (7.6) and, therefore, the linear independence of cosets (7.5). Thus, for the dimension of factorspace this yields $\dim(V/U) = n - s$, which proves the equality (7.4). The theorem is proved. $\square$

§ 8. Linear mappings.

DEFINITION 8.1. Let V and W be two linear vector spaces over a numeric field $\mathbb{K}$. A mapping $f: V \rightarrow W$ from the space V to the space W is called a *linear mapping* if the following two conditions are fulfilled:

- (1) $f(\mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{v}_1) + f(\mathbf{v}_2)$ for any two vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$;
- (2) $f(\alpha \cdot \mathbf{v}) = \alpha \cdot f(\mathbf{v})$ for any vector $\mathbf{v} \in V$ and for any number $\alpha \in \mathbb{K}$.

The relationship $f(\mathbf{0}) = \mathbf{0}$ is one of the simplest and immediate consequences of the above two properties (1) and (2) of linear mappings. Indeed, we have

$$f(\mathbf{0}) = f(\mathbf{0} + (-1) \cdot \mathbf{0}) = f(\mathbf{0}) + (-1) \cdot f(\mathbf{0}) = \mathbf{0}. \quad (8.1)$$

THEOREM 8.1. *Linear mappings possess the following three properties:*

- (1) *the identical mapping $\text{id}_V: V \rightarrow V$ of a linear vector space V onto itself is a linear mapping;*
- (2) *the composition of any two linear mappings $f: V \rightarrow W$ and $g: W \rightarrow U$ is a linear mapping $g \circ f: V \rightarrow U$;*
- (3) *if a linear mapping $f: V \rightarrow W$ is bijective, then the inverse mapping $f^{-1}: W \rightarrow V$ also is a linear mapping.*

PROOF. The linearity of the identical mapping is obvious. Indeed, here is the verification of the conditions (1) and (2) from the definition 8.1 for id_V :

$$\begin{aligned} \text{id}_V(\mathbf{v}_1 + \mathbf{v}_2) &= \mathbf{v}_1 + \mathbf{v}_2 = \text{id}_V(\mathbf{v}_1) + \text{id}_V(\mathbf{v}_2), \\ \text{id}_V(\alpha \cdot \mathbf{v}) &= \alpha \cdot \mathbf{v} = \alpha \cdot \text{id}_V(\mathbf{v}). \end{aligned}$$

Let's prove the second proposition of the theorem 8.1. Consider the composition $g \circ f$ of two linear mappings f and g . For this composition the conditions (1) and (2) from the definition 8.1 are verified as follows:

$$\begin{aligned} g \circ f(\mathbf{v}_1 + \mathbf{v}_2) &= g(f(\mathbf{v}_1 + \mathbf{v}_2)) = g(f(\mathbf{v}_1) + f(\mathbf{v}_2)) = \\ &= g(f(\mathbf{v}_1)) + g(f(\mathbf{v}_2)) = g \circ f(\mathbf{v}_1) + g \circ f(\mathbf{v}_2), \\ g \circ f(\alpha \cdot \mathbf{v}) &= g(f(\alpha \cdot \mathbf{v})) = g(\alpha \cdot f(\mathbf{v})) = \alpha \cdot g(f(\mathbf{v})) \\ &= \alpha \cdot g \circ f(\mathbf{v}). \end{aligned}$$

Now let's prove the third proposition of the theorem 8.1. Suppose that $f: V \rightarrow W$ is a bijective linear mapping. Then it possesses unique bilateral inverse mapping $f^{-1}: W \rightarrow V$ (see theorem 1.9). Let's denote

$$\begin{aligned} \mathbf{z}_1 &= f^{-1}(\mathbf{w}_1 + \mathbf{w}_2) - f^{-1}(\mathbf{w}_1) - f^{-1}(\mathbf{w}_2), \\ \mathbf{z}_2 &= f^{-1}(\alpha \cdot \mathbf{w}) - \alpha \cdot f^{-1}(\mathbf{w}). \end{aligned}$$

It is obvious that the linearity of the inverse mapping f^{-1} is equivalent to vanishing $\mathbf{z}_1$ and $\mathbf{z}_2$. Let's apply f to these vectors:

$$\begin{aligned} f(\mathbf{z}_1) &= f(f^{-1}(\mathbf{w}_1 + \mathbf{w}_2) - f^{-1}(\mathbf{w}_1) - f^{-1}(\mathbf{w}_2)) = \\ &= f(f^{-1}(\mathbf{w}_1 + \mathbf{w}_2)) - f(f^{-1}(\mathbf{w}_1)) - f(f^{-1}(\mathbf{w}_2)) = \\ &= (\mathbf{w}_1 + \mathbf{w}_2) - \mathbf{w}_1 - \mathbf{w}_2 = \mathbf{0}, \\ f(\mathbf{z}_2) &= f(f^{-1}(\alpha \cdot \mathbf{w}) - \alpha \cdot f^{-1}(\mathbf{w})) = f(f^{-1}(\alpha \cdot \mathbf{w})) - \\ &- \alpha \cdot f(f^{-1}(\mathbf{w})) = \alpha \cdot \mathbf{w} - \alpha \cdot \mathbf{w} = \mathbf{0}. \end{aligned}$$

A bijective mapping is injective. Therefore, from the equalities $f(\mathbf{z}_1) = \mathbf{0}$ and $f(\mathbf{z}_2) = \mathbf{0}$ just derived and from the equality $f(\mathbf{0}) = \mathbf{0}$ derived in (8.1) it follows that $\mathbf{z}_1 = \mathbf{z}_2 = \mathbf{0}$. The theorem is proved. $\square$

Each linear mapping $f: V \rightarrow W$ is related with two subsets: the *kernel* $\text{Ker } f \subset V$ and the *image* $\text{Im } f \subset W$. The image $\text{Im } f = f(V)$ of a linear mapping is defined in the same way as it was done for a general mapping in § 1:

$$\text{Im } f = \{\in W: \exists \mathbf{v} ((\mathbf{v} \in V) \ \& \ (f(\mathbf{v}) = \mathbf{w}))\}.$$

The kernel of a linear mapping $f: V \rightarrow W$ is the set of vectors in the space V that map to zero under the action of f :

$$\text{Ker } f = \{\mathbf{v} \in V : f(\mathbf{v}) = \mathbf{0}\}$$

THEOREM 8.2. *The kernel and the image of a linear mapping $f: V \rightarrow W$ both are subspaces in V and W respectively.*

PROOF. In order to prove this theorem we should check the conditions (1) and (2) from the definition 2.2 as applied to the subsets $\text{Ker } f \subset V$ and $\text{Im } f \subset W$.

Suppose that $\mathbf{v}_1, \mathbf{v}_2 \in \text{Ker } f$. Then $f(\mathbf{v}_1) = \mathbf{0}$ and $f(\mathbf{v}_2) = \mathbf{0}$. Suppose also that $\mathbf{v} \in \text{Ker } f$. Then $f(\mathbf{v}) = \mathbf{0}$. As a result we derive

$$\begin{aligned} f(\mathbf{v}_1 + \mathbf{v}_2) &= f(\mathbf{v}_1) + f(\mathbf{v}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0}, \\ f(\alpha \cdot \mathbf{v}) &= \alpha \cdot f(\mathbf{v}) = \alpha \cdot \mathbf{0} = \mathbf{0}. \end{aligned}$$

Hence, $\mathbf{v}_1 + \mathbf{v}_2 \in \text{Ker } f$ and $\alpha \cdot \mathbf{v} \in \text{Ker } f$. This proves the proposition of the theorem concerning the kernel $\text{Ker } f$.

Let $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w} \in \text{Im } f$. Then there are three vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}$ in V such that $f(\mathbf{v}_1) = \mathbf{w}_1$, $f(\mathbf{v}_2) = \mathbf{w}_2$, and $f(\mathbf{v}) = \mathbf{w}$. Hence, we have

$$\begin{aligned} \mathbf{w}_1 + \mathbf{w}_2 &= f(\mathbf{v}_1) + f(\mathbf{v}_2) = f(\mathbf{v}_1 + \mathbf{v}_2), \\ \alpha \cdot \mathbf{w} &= \alpha \cdot f(\mathbf{v}) = f(\alpha \cdot \mathbf{v}). \end{aligned}$$

This meant that $\mathbf{w}_1 + \mathbf{w}_2 \in \text{Im } f$ and $\alpha \cdot \mathbf{w} \in \text{Im } f$. The theorem is proved. $\square$

Remember that, according to the theorem 1.2, a linear mapping $f: V \rightarrow W$ is surjective if and only if $\text{Im } f = W$. There is a similar proposition for $\text{Ker } f$.

THEOREM 8.3. *A linear mapping $f: V \rightarrow W$ is injective if and only if its kernel is zero, i. e. $\text{Ker } f = \{\mathbf{0}\}$.*

PROOF. Let f be injective and let $\mathbf{v} \in \text{Ker } f$. Then $f(\mathbf{0}) = \mathbf{0}$ and $f(\mathbf{v}) = \mathbf{0}$. But if $\mathbf{v} \neq \mathbf{0}$, then due to injectivity of f it would be $f(\mathbf{v}) \neq f(\mathbf{0})$. Hence, $\mathbf{v} = \mathbf{0}$. This means that the kernel of f consists of the only one element: $\text{Ker } f = \{\mathbf{0}\}$.

Now conversely, suppose that $\text{Ker } f = \{\mathbf{0}\}$. Let's consider two different vectors $\mathbf{v}_1 \neq \mathbf{v}_2$ in V . Then $\mathbf{v}_1 - \mathbf{v}_2 \neq \mathbf{0}$ and $\mathbf{v}_1 - \mathbf{v}_2 \notin \text{Ker } f$. Therefore, $f(\mathbf{v}_1 - \mathbf{v}_2) \neq \mathbf{0}$. Applying the linearity of f , from this inequality we derive $f(\mathbf{v}_1) - f(\mathbf{v}_2) \neq \mathbf{0}$, i. e. $f(\mathbf{v}_1) \neq f(\mathbf{v}_2)$. Hence, f is an injective mapping. The theorem is proved. $\square$

The following theorem is known as the theorem on the linear independence of preimages. Here is its statement.

THEOREM 8.4. *Let $f: V \rightarrow W$ be a linear mapping and let $\mathbf{v}_1, \dots, \mathbf{v}_s$ be some vectors of a linear vector space V such that their images $f(\mathbf{v}_1), \dots, f(\mathbf{v}_s)$ in W are linearly independent. Then the vectors $\mathbf{v}_1, \dots, \mathbf{v}_s$ themselves are also linearly independent.*

PROOF. In order to prove the theorem let's consider a linear combination of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_s$ being equal to zero:

$$\alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_s \cdot \mathbf{v}_s = \mathbf{0}.$$

Applying f to both sides of this equality and using the fact that f is a linear mapping, we obtain quite similar equality for the images

$$\alpha_1 \cdot f(\mathbf{v}_1) + \dots + \alpha_s \cdot f(\mathbf{v}_s) = \mathbf{0}.$$

However, these images $f(\mathbf{v}_1), \dots, f(\mathbf{v}_s)$ are linearly independent. Hence, all coefficients in the above linear combination are equal to zero: $\alpha_1 = \dots = \alpha_s = 0$.

Then the initial linear combination is also necessarily trivial. This proves that the vectors $\mathbf{v}_1, \dots, \mathbf{v}_s$ are linearly independent. $\square$

A linear vector space is a set. But it is not simply a set — it is a structured set. It is equipped with algebraic operations satisfying the axioms (1)-(8). Linear mappings are those being concordant with the structures of a linear vector space in the spaces they are acting from and to. In algebra such mappings concordant with algebraic structures are called *morphisms*. So, in algebraic terminology, linear mappings are morphisms of linear vector spaces.

DEFINITION 8.2. Two linear vector spaces V and W are called *isomorphic* if there is a bijective linear mapping $f: V \rightarrow W$ binding them.

The first example of an isomorphism of linear vector spaces is the mapping $\psi: V \rightarrow \mathbb{K}^n$ in (5.4). Because of the existence of such mapping we can formulate the following theorem.

THEOREM 8.5. *Any n -dimensional linear vector space V is isomorphic to the arithmetic linear vector space $\mathbb{K}^n$.*

Isomorphic linear vector spaces have many common features. Often they can be treated as undistinguishable. In particular, we have the following fact.

THEOREM 8.6. *If a linear vector space V is isomorphic to a finite-dimensional vector space W , then V is also finite-dimensional and the dimensions of these two spaces do coincide: $\dim V = \dim W$.*

PROOF. Let $f: V \rightarrow W$ be an isomorphism of spaces V and W . Assume for the sake of certainty that $\dim W = n$ and choose a basis $\mathbf{h}_1, \dots, \mathbf{h}_n$ in W . By means of inverse mapping $f^{-1}: W \rightarrow V$ we define the vectors $\mathbf{e}_i = f^{-1}(\mathbf{h}_i)$, $i = 1, \dots, n$. Let $\mathbf{v}$ be an arbitrary vector of V . Let's map it with the use of f into the space W and then expand in the basis:

$$f(\mathbf{v}) = \alpha_1 \cdot \mathbf{h}_1 + \dots + \alpha_n \cdot \mathbf{h}_n.$$

Applying the inverse mapping f^{-1} to both sides of this equality, due to the linearity of f^{-1} we get the expansion

$$\mathbf{v} = \alpha_1 \cdot \mathbf{e}_1 + \dots + \alpha_n \cdot \mathbf{e}_n.$$

From this expansion we derive that $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a finite spanning system in V . The finite dimensionality of V is proved. The linear independence of $\mathbf{e}_1, \dots, \mathbf{e}_n$ follows from the theorem 8.4 on the linear independence of preimages. Hence, $\mathbf{e}_1, \dots, \mathbf{e}_n$ is a basis in V and $\dim V = n = \dim W$. The theorem is proved. $\square$

§ 9. The matrix of a linear mapping.

Let $f: V \rightarrow W$ be a linear mapping from n -dimensional vector space V to m -dimensional vector space W . Let's choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in V and a basis $\mathbf{h}_1, \dots, \mathbf{h}_m$ in W . Then consider the images of basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ in W and

Remember that when composing a column vector of the coordinates of a vector $\mathbf{x}$, we negotiated to understand this procedure as a linear mapping $\psi: V \rightarrow \mathbb{K}^n$ (see formulas (5.4) and the theorem 8.5). Denote by $\tilde{\psi}: W \rightarrow \mathbb{K}^m$ the analogous mapping for a vector $\mathbf{y}$ in W . Then the matrix relationship (9.5) can be treated as a mapping $F: \mathbb{K}^n \rightarrow \mathbb{K}^m$. These three mappings $\psi, \tilde{\psi}, F$ and the initial mapping f can be written in a diagram:

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ \psi \downarrow & & \downarrow \tilde{\psi} \\ \mathbb{K}^n & \xrightarrow{F} & \mathbb{K}^m \end{array} \quad (9.6)$$

Such diagrams are called *commutative diagrams* if the compositions of mappings «when passing along arrows» from any node to any other node do not depend on a particular path connecting these two nodes. When applied to the diagram (9.6), the commutativity means $\tilde{\psi} \circ f = F \circ \psi$. Due to the bijectivity of linear mappings ψ and $\tilde{\psi}$ the condition of commutativity of the diagram (9.6) can be written as

$$F = \tilde{\psi} \circ f \circ \psi^{-1}, \quad f = \tilde{\psi}^{-1} \circ F \circ \psi. \quad (9.7)$$

The reader can easily check that the relationships (9.7) are fulfilled due to the way the matrix F is constructed. Hence, the diagram (9.6) is commutative.

Now let's look at the relationships (9.7) from a little bit other point of view. Let V and W be two spaces of the dimensions n and m respectively. Suppose that we have an arbitrary $m \times n$ matrix F . Then the relationship (9.5) determines a linear mapping $F: \mathbb{K}^n \rightarrow \mathbb{K}^m$. Choosing bases $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $\mathbf{h}_1, \dots, \mathbf{h}_m$ in V and W we can use the second relationship (9.7) in order to define the linear mapping $f: V \rightarrow W$. The matrix of this mapping in the bases $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $\mathbf{h}_1, \dots, \mathbf{h}_m$ coincides with F exactly. Thus, we have proved the following theorem.

THEOREM 9.1. *Any rectangular $m \times n$ matrix F can be constructed as a matrix of a linear mapping $f: V \rightarrow W$ from n -dimensional vector space V to m -dimensional vector space W in some pair of bases in these spaces.*

A more straightforward way of proving the theorem 9.1 than we considered above can be based on the following theorem.

THEOREM 9.2. *For any basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in n -dimensional vector space V and for any set of n vectors $\mathbf{w}_1, \dots, \mathbf{w}_n$ in another vector space W there is a linear mapping $f: V \rightarrow W$ such that $f(\mathbf{e}_i) = \mathbf{w}_i$ for $i = 1, \dots, n$.*

PROOF. Once the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in V is chosen, this defines the mapping $\psi: V \rightarrow \mathbb{K}^n$ (see (5.4)). In order to construct the required mapping f we define a mapping $\varphi: \mathbb{K}^n \rightarrow W$ by the following relationship:

$$\varphi: \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} \mapsto x^1 \cdot \mathbf{w}_1 + \dots + x^n \cdot \mathbf{w}_n.$$

Now it is easy to verify that the required mapping is the composition $f = \varphi \circ \psi$. $\square$

Let's return to initial situation. Suppose that we have a mapping $f: V \rightarrow W$ that determines a matrix F upon choosing two bases $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $\mathbf{h}_1, \dots, \mathbf{h}_m$ in V and W respectively. The matrix F essentially depends on the choice of bases. In order to describe this dependence we consider four bases — two bases in V and other two bases in W . Suppose that S and P are direct transition matrices for that pairs of bases. Their components are defined as follows:

$$\tilde{\mathbf{e}}_k = \sum_{j=1}^n S_k^j \cdot \mathbf{e}_j, \quad \tilde{\mathbf{h}}_r = \sum_{i=1}^m P_r^i \cdot \mathbf{h}_i.$$

The inverse transition matrices $T = S^{-1}$ and $Q = P^{-1}$ are defined similarly:

$$\mathbf{e}_j = \sum_{k=1}^n T_j^k \cdot \tilde{\mathbf{e}}_k, \quad \mathbf{h}_i = \sum_{r=1}^m Q_i^r \cdot \tilde{\mathbf{h}}_r.$$

We use these relationships and the above relationships (9.3) in order to carry out the following calculations for the vector $f(\tilde{\mathbf{e}}_k)$:

$$\begin{aligned} f(\tilde{\mathbf{e}}_k) &= \sum_{j=1}^n S_k^j \cdot f(\mathbf{e}_j) = \sum_{j=1}^n S_k^j \cdot \left(\sum_{i=1}^m F_j^i \cdot \mathbf{h}_i \right) = \\ &= \sum_{j=1}^n S_k^j \cdot \left(\sum_{i=1}^m F_j^i \cdot \left(\sum_{r=1}^m Q_i^r \cdot \tilde{\mathbf{h}}_r \right) \right). \end{aligned}$$

Upon changing the order of summations this result is written as

$$f(\tilde{\mathbf{e}}_k) = \sum_{r=1}^m \left(\sum_{i=1}^m \sum_{j=1}^n Q_i^r F_j^i S_k^j \right) \cdot \tilde{\mathbf{h}}_r.$$

The double sums in round brackets are the coefficients of the expansion of the vector $f(\tilde{\mathbf{e}}_k)$ in the basis $\tilde{\mathbf{h}}_1, \dots, \tilde{\mathbf{h}}_m$. They determine the matrix of the linear mapping f in wavy bases $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ and $\tilde{\mathbf{h}}_1, \dots, \tilde{\mathbf{h}}_m$:

$$\tilde{F}_k^r = \sum_{i=1}^m \sum_{j=1}^n Q_i^r F_j^i S_k^j. \quad (9.8)$$

In a similar way one can derive the converse relationship expressing F through $\tilde{F}$:

$$F_j^i = \sum_{r=1}^m \sum_{k=1}^n P_r^i \tilde{F}_k^r T_j^k. \quad (9.9)$$

The relationships (9.8) and (9.9) are called the *transformation formulas for the matrix of a linear mapping under a change of bases*. They can be written as

$$\tilde{F} = P^{-1} F S, \quad F = P \tilde{F} S^{-1}. \quad (9.10)$$

This is the matrix form of the relationships (9.8) and (9.9).

The transformation formulas like (9.10) lead us to the broad class of problems of «bringing to a canonic form». In our particular case a change of bases in the

spaces V and W changes the matrix of the linear mapping $f: V \rightarrow W$. The problem of bringing to a canonic form in this case consists in finding the optimal choice of bases, where the matrix F has the most simple (canonic) form. The following theorem solving this particular problem is known as the theorem on bringing to the *almost diagonal* form.

THEOREM 9.3. *Let $f: V \rightarrow W$ be some nonzero linear mapping from n -dimensional vector space V to m -dimensional vector space W . Then there is a choice of bases in V and W such that the matrix F of this mapping has the following almost diagonal form:*

$$F = \left\| \begin{array}{cccccccc} \overbrace{1 & 0 & \dots & 0}^s & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \end{array} \right\|_s \quad (9.11)$$

PROOF. The purely zero mapping $0: V \rightarrow W$ maps each vector of the space V to zero vector in W . The matrix of such mapping consists of zeros only. There is no need to formulate the problem of bringing it to a canonic form.

Let $f: V \rightarrow W$ be a nonzero linear mapping. The integer number $s = \dim(\text{Im } f)$ is called the *rank* of the mapping f . The rank of a nonzero mapping is not equal to zero. We begin constructing a canonic base in W by choosing a base $\mathbf{h}_1, \dots, \mathbf{h}_s$ in the image space $\text{Im } f$. For each basis vector $\mathbf{h}_i \in \text{Im } f$ there is a vector $\mathbf{e}_i \in V$ such that $f(\mathbf{e}_i) = \mathbf{h}_i$, $i = 1, \dots, s$. These vectors $\mathbf{e}_1, \dots, \mathbf{e}_s$ are linearly independent due to the theorem 8.4. Let $r = \dim(\text{Ker } f)$. We choose a basis in $\text{Ker } f$ and denote the basis vectors by $\mathbf{e}_{s+1}, \dots, \mathbf{e}_{s+r}$. Then we consider the vectors

$$\mathbf{e}_1, \dots, \mathbf{e}_s, \mathbf{e}_{s+1}, \dots, \mathbf{e}_{s+r} \quad (9.12)$$

and prove that they form a basis in V . For this purpose we use the theorem 4.6.

Let's begin with checking the condition (1) in the theorem 4.6 for the vectors (9.12). In order to prove the linear independence of these vectors we consider a linear combination of them being equal to zero:

$$\alpha_1 \cdot \mathbf{e}_1 + \dots + \alpha_s \cdot \mathbf{e}_s + \alpha_{s+1} \cdot \mathbf{e}_{s+1} + \dots + \alpha_{s+r} \cdot \mathbf{e}_{s+r} = \mathbf{0}. \quad (9.13)$$

Let's apply the mapping f to both sides of the equality (9.13) and take into account that $f(\mathbf{e}_i) = \mathbf{h}_i$ for $i = 1, \dots, s$. Other vectors belong to the kernel of the mapping f , therefore, $f(\mathbf{e}_{s+i}) = \mathbf{0}$ for $i = 1, \dots, r$. Then from (9.13) we derive

$$\alpha_1 \cdot \mathbf{h}_1 + \dots + \alpha_s \cdot \mathbf{h}_s = \mathbf{0}.$$

The vectors $\mathbf{h}_1, \dots, \mathbf{h}_s$ form a basis in $\text{Im } f$. They are linearly independent.

Hence, $\alpha_1 = \dots = \alpha_s = 0$. Taking into account this fact, we reduce (9.13) to

$$\alpha_{s+1} \cdot \mathbf{e}_{s+1} + \dots + \alpha_{s+r} \cdot \mathbf{e}_{s+r} = \mathbf{0}.$$

The vectors $\mathbf{e}_{s+1}, \dots, \mathbf{e}_{s+r}$ form a basis in $\text{Ker } f$. They are linearly independent, therefore, $\alpha_{s+1} = \dots = \alpha_{s+r} = 0$. As a result we have proved that all coefficients of the linear combination (9.13) are necessarily zero. Hence, the vectors (9.12) are linearly independent.

Now let's check the second condition of the theorem 4.6 for the vectors (9.12). Assume that $\mathbf{v}$ is an arbitrary vector in V . Then $f(\mathbf{v})$ belongs to $\text{Im } f$. Let's expand $f(\mathbf{v})$ in the basis $\mathbf{h}_1, \dots, \mathbf{h}_s$:

$$f(\mathbf{v}) = \beta_1 \cdot \mathbf{h}_1 + \dots + \beta_s \cdot \mathbf{h}_s. \quad (9.14)$$

Remember that $f(\mathbf{e}_i) = \mathbf{h}_i$ for $i = 1, \dots, s$. Then from (9.14) we derive

$$\begin{aligned} \mathbf{0} &= f(\mathbf{v}) - \beta_1 \cdot f(\mathbf{e}_1) - \dots - \beta_s \cdot f(\mathbf{e}_s) = \\ &= f(\mathbf{v} - \beta_1 \cdot \mathbf{e}_1 - \dots - \beta_s \cdot \mathbf{e}_s). \end{aligned} \quad (9.15)$$

Let's denote $\tilde{\mathbf{v}} = \mathbf{v} - \beta_1 \cdot \mathbf{e}_1 - \dots - \beta_s \cdot \mathbf{e}_s$. From (9.15) we derive $f(\tilde{\mathbf{v}}) = \mathbf{0}$ for this vector $\tilde{\mathbf{v}}$. Hence, $\tilde{\mathbf{v}} \in \text{Ker } f$. Let's expand $\tilde{\mathbf{v}}$ in the basis of $\text{Ker } f$:

$$\tilde{\mathbf{v}} = \beta_{s+1} \cdot \mathbf{e}_{s+1} + \dots + \beta_{s+r} \cdot \mathbf{e}_{s+r}.$$

From the formula $\tilde{\mathbf{v}} = \mathbf{v} - \beta_1 \cdot \mathbf{e}_1 - \dots - \beta_s \cdot \mathbf{e}_s$ and the above expansion we get

$$\mathbf{v} = \beta_1 \cdot \mathbf{e}_1 + \dots + \beta_s \cdot \mathbf{e}_s + \beta_{s+1} \cdot \mathbf{e}_{s+1} + \dots + \beta_{s+r} \cdot \mathbf{e}_{s+r}.$$

This means that the vectors (9.12) form a spanning system in V . The condition (2) of the theorem 4.6 for them is also fulfilled. Thus, the vectors (9.12) form a basis in V . This yields the equality

$$\dim V = s + r. \quad (9.16)$$

In order to complete the proof of the theorem we need to complete the basis $\mathbf{h}_1, \dots, \mathbf{h}_s$ of $\text{Im } f$ up to a basis $\mathbf{h}_1, \dots, \mathbf{h}_s, \mathbf{h}_{s+1}, \dots, \mathbf{h}_m$ in the space W . For the vector $f(\mathbf{e}_j)$ with $j = 1, \dots, s$ we have the expansion

$$f(\mathbf{e}_j) = \mathbf{h}_j = \sum_{i=1}^s \delta_j^i \cdot \mathbf{h}_i + \sum_{i=s+1}^m 0 \cdot \mathbf{h}_i.$$

If $j = s+1, \dots, s+r$, the expansion for $f(\mathbf{e}_j)$ is purely zero:

$$f(\mathbf{e}_j) = \mathbf{0} = \sum_{i=1}^s 0 \cdot \mathbf{h}_i + \sum_{i=s+1}^m 0 \cdot \mathbf{h}_i.$$

Due to these expansions the matrix of the mapping f in the bases that we have constructed above has the required almost diagonal form (9.10). $\square$

In proving this theorem we have proved simultaneously the next one.

THEOREM 9.4. *Let $f: V \rightarrow W$ be a linear mapping from n -dimensional space V to an arbitrary linear vector space W . Then*

$$\dim(\text{Ker } f) + \dim(\text{Im } f) = \dim V. \quad (9.17)$$

This theorem 9.4 is known as the theorem *on the sum of dimensions of the kernel and the image* of a linear mapping. The proposition of the theorem in the form of the relationship (9.17) immediately follows from (9.16).

§ 10. Algebraic operations with mappings. The space of homomorphisms $\text{Hom}(V, W)$.

DEFINITION 10.1. Let V and W be two linear vector spaces and let $f: V \rightarrow W$ and $g: V \rightarrow W$ be two linear mappings from V to W . The linear mapping $h: V \rightarrow W$ defined by the relationship $h(\mathbf{v}) = f(\mathbf{v}) + g(\mathbf{v})$, where $\mathbf{v}$ is an arbitrary vector of V , is called the *sum* of the mappings f and g .

DEFINITION 10.2. Let V and W be two linear vector spaces over a numeric field $\mathbb{K}$ and let $f: V \rightarrow W$ be a linear mapping from V to W . The linear mapping $h: V \rightarrow W$ defined by the relationship $h(\mathbf{v}) = \alpha \cdot f(\mathbf{v})$, where $\mathbf{v}$ is an arbitrary vector of V , is called the *product* of the number $\alpha \in \mathbb{K}$ and the mapping f .

The algebraic operations introduced by the definitions 10.1 and 10.2 are called *pointwise addition* and *pointwise multiplication by a number*. Indeed, they are calculated «pointwise» by adding the values of the initial mappings and by multiplying them by a number for each specific argument $\mathbf{v} \in V$. These operations are denoted by the same signs as the corresponding operations with vectors: $h = f + g$ and $h = \alpha \cdot f$. The writing $(f + g)(\mathbf{v})$ is understood as the sum of mappings applied to the vector $\mathbf{v}$. Another writing $f(\mathbf{v}) + g(\mathbf{v})$ denotes the sum of the results of applying f and g to $\mathbf{v}$ separately. Though the results of these calculations do coincide, their meanings are different. In a similar way one should distinguish the meanings of left and right sides of the following equality:

$$(\alpha \cdot f)(\mathbf{v}) = \alpha \cdot f(\mathbf{v}).$$

Let's denote by $\text{Map}(V, W)$ the set of all mappings from the space V to the space W . Sometimes this set is denoted by W^V .

THEOREM 10.1. *Let V and W be two linear spaces over a numeric field $\mathbb{K}$. Then the set of mappings $\text{Map}(V, W)$ equipped with the operations of pointwise addition and pointwise multiplication by numbers fits the definition of a linear vector space over the numeric field $\mathbb{K}$.*

PROOF. Let's verify the axioms of a linear vector space for the set of mappings $\text{Map}(V, W)$. In the case of the first axiom we should verify the coincidence of the mappings $f + g$ and $g + f$. Remember that the coincidence of two mappings is equivalent to the coincidence of their values when applied to an arbitrary vector $v \in V$. The following calculations establish the latter coincidence:

$$(f + g)(\mathbf{v}) = f(\mathbf{v}) + g(\mathbf{v}) = g(\mathbf{v}) + f(\mathbf{v}) = (g + f)(\mathbf{v}).$$

As we see in the above calculations, the equality $f + g = g + f$ follows from the commutativity axiom for the addition of vectors in W due to pointwise nature of the addition of mappings. The same arguments are applicable when verifying the axioms (2), (5), and (6) for the algebraic operations with mappings:

$$\begin{aligned}
 ((f + g) + h)(\mathbf{v}) &= (f + g)(\mathbf{v}) + h(\mathbf{v}) = (f(\mathbf{v}) + g(\mathbf{v})) + h(\mathbf{v}) = \\
 &= f(\mathbf{v}) + (g(\mathbf{v}) + h(\mathbf{v})) = f(\mathbf{v}) + (g + h)(\mathbf{v}) = (f + (g + h))(\mathbf{v}) \\
 (\alpha \cdot (f + g))(\mathbf{v}) &= \alpha \cdot (f + g)(\mathbf{v}) = \alpha \cdot (f(\mathbf{v}) + g(\mathbf{v})) = \\
 &= \alpha \cdot f(\mathbf{v}) + \alpha \cdot g(\mathbf{v}) = (\alpha \cdot f)(\mathbf{v}) + (\alpha \cdot g)(\mathbf{v}) = (\alpha \cdot f + \alpha \cdot g)(\mathbf{v}) \\
 ((\alpha + \beta) \cdot f)(\mathbf{v}) &= (\alpha + \beta) \cdot f(\mathbf{v}) = \alpha \cdot f(\mathbf{v}) + \beta \cdot f(\mathbf{v}) = \\
 &= (\alpha \cdot f)(\mathbf{v}) + (\beta \cdot f)(\mathbf{v}) = (\alpha \cdot f + \beta \cdot f)(\mathbf{v})
 \end{aligned}$$

For the axioms (7) these calculations look like

$$(\alpha \cdot (\beta \cdot f))(\mathbf{v}) = \alpha \cdot (\beta \cdot f)(\mathbf{v}) = \alpha \cdot (\beta \cdot f(\mathbf{v})) = (\alpha\beta) \cdot f(\mathbf{v}) = ((\alpha\beta) \cdot f)(\mathbf{v}).$$

In the case of the axiom (8) the calculations are even more simple:

$$(1 \cdot f)(\mathbf{v}) = 1 \cdot f(\mathbf{v}) = f(\mathbf{v}).$$

Now let's consider the rest axioms (3) and (4). The zero mapping is the best pretender for the role of zero element in the space $\text{Map}(V, W)$, it maps each vector $\mathbf{v} \in V$ to zero vector of the space W . For this mapping we have

$$(f + 0)(\mathbf{v}) = f(\mathbf{v}) + 0(\mathbf{v}) = f(\mathbf{v}) + \mathbf{0} = f(\mathbf{v}).$$

As we see, the axiom (3) in $\text{Map}(V, W)$ is fulfilled.

Suppose that $f \in \text{Map}(V, W)$. We define the opposite mapping f' for f as follows: $f' = (-1) \cdot f$. Then we have

$$\begin{aligned}
 (f + f')(\mathbf{v}) &= (f + (-1) \cdot f)(\mathbf{v}) = f(\mathbf{v}) + \\
 &+ ((-1) \cdot f)(\mathbf{v}) = f(\mathbf{v}) + (-1) \cdot f(\mathbf{v}) = \mathbf{0} = 0(\mathbf{v}).
 \end{aligned}$$

The axiom (4) in $\text{Map}(V, W)$ is also fulfilled. This completes the proof of the theorem 10.1. $\square$

In typical situation the space $\text{Map}(V, W)$ is very large. Even for the finite-dimensional spaces V and W usually it is an infinite-dimensional space. In linear algebra the much smaller subset of $\text{Map}(V, W)$ is studied. This is the set of all linear mappings from V to W . It is denoted $\text{Hom}(V, W)$ and is called the *set of homomorphisms*. The following two theorems show that $\text{Hom}(V, W)$ is closed with respect to algebraic operations in $\text{Map}(V, W)$. Therefore, we can say that $\text{Hom}(V, W)$ is the *space of homomorphisms*.

THEOREM 10.2. *The pointwise sum of two linear mappings $f : V \rightarrow W$ and $g : V \rightarrow W$ is a linear mapping from the space V to the space W .*

THEOREM 10.3. *The pointwise product of a linear mapping $f: V \rightarrow W$ by a number $\alpha \in \mathbb{K}$ is a linear mapping from the space V to the space W .*

PROOF. Let $h = f + g$ be the sum of two linear mappings f and g . The following calculations prove the linearity of the mapping h :

$$\begin{aligned} h(\mathbf{v}_1 + \mathbf{v}_2) &= f(\mathbf{v}_1 + \mathbf{v}_2) + g(\mathbf{v}_1 + \mathbf{v}_2) = (f(\mathbf{v}_1) + \\ &\quad + f(\mathbf{v}_2)) + (g(\mathbf{v}_1) + g(\mathbf{v}_2)) = (f(\mathbf{v}_1) + \\ &\quad + g(\mathbf{v}_1)) + (f(\mathbf{v}_2) + g(\mathbf{v}_2)) = h(\mathbf{v}_1) + h(\mathbf{v}_2), \\ h(\beta \cdot \mathbf{v}) &= f(\beta \cdot \mathbf{v}) + g(\beta \cdot \mathbf{v}) = \beta \cdot f(\mathbf{v}) + \\ &\quad + \beta \cdot g(\mathbf{v}) = \beta \cdot (f(\mathbf{v}) + g(\mathbf{v})) = \beta \cdot h(\mathbf{v}). \end{aligned}$$

Now let's consider the product of the mapping f and the number α . Let's denote it by h , i.e. let's denote $h = \alpha \cdot f$. Then the following calculations

$$\begin{aligned} h(\mathbf{v}_1 + \mathbf{v}_2) &= \alpha \cdot f(\mathbf{v}_1 + \mathbf{v}_2) = \alpha \cdot (f(\mathbf{v}_1) + \\ &\quad + f(\mathbf{v}_2)) = \alpha \cdot f(\mathbf{v}_1) + \alpha \cdot f(\mathbf{v}_2) = h(\mathbf{v}_1) + h(\mathbf{v}_2), \\ h(\beta \cdot \mathbf{v}) &= \alpha \cdot f(\beta \cdot \mathbf{v}) = \alpha \cdot (\beta \cdot f(\mathbf{v})) = \\ &= (\alpha\beta) \cdot f(\mathbf{v}) = (\beta\alpha) \cdot f(\mathbf{v}) = \beta \cdot (\alpha \cdot f(\mathbf{v})) = \beta \cdot h(\mathbf{v}). \end{aligned}$$

prove the linearity of the mapping h and thus complete the proofs of both theorems 10.2 and 10.3. $\square$

The space of homomorphisms $\text{Hom}(V, W)$ is a subspace in the space of all mappings $\text{Map}(V, W)$. It is much smaller and it consists of objects which are in the scope of linear algebra. For finite-dimensional spaces V and W the space of homomorphisms $\text{Hom}(V, W)$ is also finite-dimensional. This is the result of the following theorem.

THEOREM 10.4. *For finite-dimensional spaces V and W the space of homomorphisms $\text{Hom}(V, W)$ is also finite-dimensional. Its dimension is given by formula*

$$\dim(\text{Hom}(V, W)) = \dim(V) \cdot \dim(W). \quad (10.1)$$

PROOF. Let $\dim V = n$ and $\dim W = m$. We choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in the space V and another basis $\mathbf{h}_1, \dots, \mathbf{h}_m$ in the space W . Let $1 \leq i \leq n$ and $1 \leq j \leq m$. For each fixed pair of indices i, j within the above ranges we consider the following set of n vectors in the space W :

$$\mathbf{w}_1 = \mathbf{0}, \dots, \mathbf{w}_{i-1} = \mathbf{0}, \mathbf{w}_i = \mathbf{h}_j, \mathbf{w}_{i+1} = \mathbf{0}, \dots, \mathbf{w}_n = \mathbf{0}.$$

All vectors in this set are equal to zero, except for the i -th vector $\mathbf{w}_i$ which is equal to j -th basis vector $\mathbf{h}_j$. Now we apply the theorem 9.2 to the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in V and to the set of vector $\mathbf{w}_1, \dots, \mathbf{w}_n$. This defines the linear mapping $E_j^i: V \rightarrow W$ such that $E_j^i(\mathbf{e}_s) = \mathbf{w}_s$ for all $s = 1, \dots, n$. We write this fact as

$$E_j^i(\mathbf{e}_s) = \delta_s^i \cdot \mathbf{h}_j, \quad (10.2)$$

where δ_s^i is the Kronecker symbol. As a result we have constructed $n \cdot m$ mappings E_j^i satisfying the relationships (10.2):

$$E_j^i: V \rightarrow W, \text{ where } 1 \leq i \leq n, 1 \leq j \leq m. \quad (10.3)$$

Now we show that the mapping (10.3) span the space of homomorphisms $\text{Hom}(V, W)$. For this purpose we take a linear mapping $f \in \text{Hom}(V, W)$. Suppose that F is its matrix in the pair of bases $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $\mathbf{h}_1, \dots, \mathbf{h}_m$. Denote by F_i^j the elements of this matrix. Then the result of applying f to an arbitrary vector $\mathbf{v} \in V$ is determined by coordinates of this vector according to the formula

$$f(\mathbf{v}) = \sum_{i=1}^n v^i \cdot f(\mathbf{e}_i) = \sum_{i=1}^n \sum_{j=1}^m (F_i^j v^i) \cdot \mathbf{h}_j. \quad (10.4)$$

Applying E_j^i to the same vector $\mathbf{v}$ and taking into account (10.2), we derive

$$E_j^i(\mathbf{v}) = \sum_{s=1}^n v^s \cdot E_j^i(\mathbf{e}_s) = \sum_{s=1}^n (v^s \delta_s^i) \cdot \mathbf{h}_j = v^i \cdot \mathbf{h}_j. \quad (10.5)$$

Now, comparing the relationships (10.4) and (10.5), we find

$$f(\mathbf{v}) = \sum_{i=1}^n \sum_{j=1}^m F_i^j \cdot E_j^i(\mathbf{v}).$$

Since $\mathbf{v}$ is an arbitrary vector of the space V , this formula means that f is a linear combination of the mappings (10.3):

$$f = \sum_{i=1}^n \sum_{j=1}^m F_i^j \cdot E_j^i.$$

Hence, the mappings (10.3) span the space of homomorphisms $\text{Hom}(V, W)$. This proves the finite-dimensionality of the space $\text{Hom}(V, W)$.

In order to calculate the dimension of $\text{Hom}(V, W)$ we shall prove that the mappings (10.3) are linearly independent. Let's consider a linear combination of these mappings, which is equal to zero:

$$\sum_{i=1}^n \sum_{j=1}^m \gamma_i^j \cdot E_j^i = 0. \quad (10.6)$$

Both left and right hand sides of the equality (10.6) represent the zero mapping $0: V \rightarrow W$. Let's apply this mapping to the basis vector $\mathbf{e}_s$. Then

$$\sum_{i=1}^n \sum_{j=1}^m \gamma_i^j \cdot E_j^i(\mathbf{e}_s) = \sum_{i=1}^n \sum_{j=1}^m (\gamma_i^j \delta_s^i) \cdot \mathbf{h}_j = 0.$$

The sum in the index i can be calculated explicitly. As a result we get the linear combinations of basis vectors in W , which are equal to zero:

$$\sum_{j=1}^m \gamma_i^j \cdot \mathbf{h}_j = 0.$$

Due to the linear independence of the vectors $\mathbf{h}_1, \dots, \mathbf{h}_m$ we derive $\gamma_i^j = 0$. This means that the linear combination (10.6) is necessarily trivial. Hence, the mappings (10.3) are linearly independent. They form a basis in $\text{Hom}(V, W)$. Now, by counting these mappings we find that the required formula (10.2) is valid. $\square$

The meaning of the above theorem becomes transparent in terms of the matrices of linear mappings. Indeed, upon choosing the bases in V and W the linear mappings from $\text{Hom}(V, W)$ are represented by rectangular $m \times n$ matrices. The sum of mappings corresponds to the sum of matrices, and the product of a mapping by a number corresponds to the product of the matrix by that number. Note that rectangular $m \times n$ matrices form a linear vector space isomorphic to the arithmetic linear vector space $\mathbb{K}^{mn}$. This space is denoted as $\mathbb{K}^{m \times n}$. So, the choice of bases in V and W defines an isomorphism of $\text{Hom}(V, W)$ and $\mathbb{K}^{m \times n}$.

CHAPTER II

LINEAR OPERATORS.

§ 1. Linear operators. The algebra of endomorphisms End(V) and the group of automorphisms Aut(V).

A linear mapping $f: V \rightarrow V$ acting from a linear vector space V to the same vector space V is called a *linear operator*¹. Linear operators are special forms of linear mappings. Therefore, we can apply to them all results of previous chapter. However, the less generality the more specific features. Therefore, the theory of linear operators appears to be more rich and more complicated than the theory of linear mappings. It contains not only the strengthening of previous theorems for this particular case, but a class of problems that cannot be formulated for the case of general linear mappings.

Let's consider the space of homomorphisms $\text{Hom}(V, W)$. If $W = V$, this space is called the *space of endomorphisms* $\text{End}(V) = \text{Hom}(V, V)$. It consists of linear operators $f: V \rightarrow V$ which are also called *endomorphisms* of the space V . Unlike the space of homomorphisms $\text{Hom}(V, W)$, the space of endomorphisms $\text{End}(V)$ is equipped with the additional binary algebraic operation. Indeed, if we have two linear operators $f, g \in \text{End}(V)$, we can not only add them and multiply them by numbers, but we can also construct two compositions $f \circ g \in \text{End}(V)$ and $g \circ f \in \text{End}(V)$.

THEOREM 1.1. *Let $\text{End}(V)$ be the space of endomorphisms of a linear vector space V . Here, apart from the axioms (1)-(8) of a linear vector space, the following relationships are fulfilled:*

$$\begin{array}{ll} (9) \quad (f + g) \circ h = f \circ h + g \circ h; & (11) \quad f \circ (g + h) = f \circ g + f \circ h; \\ (10) \quad (\alpha \cdot f) \circ h = \alpha \cdot (f \circ h); & (12) \quad f \circ (\alpha \cdot g) = \alpha \cdot (f \circ g); \end{array}$$

PROOF. Each of the equalities (9)-(12) is an operator equality. As we know, the equality of two operators means that these operators yield the same result when applied to an arbitrary vector $v \in V$:

$$\begin{aligned} ((f + g) \circ h)(\mathbf{v}) &= (f + g)(h(\mathbf{v})) = f(h(\mathbf{v})) + g(h(\mathbf{v})) = \\ &= (f \circ h)(\mathbf{v}) + (g \circ h)(\mathbf{v}) = (f \circ h + g \circ h)(\mathbf{v}) \end{aligned}$$

$$\begin{aligned} ((\alpha \cdot f) \circ h)(\mathbf{v}) &= (\alpha \cdot f)(h(\mathbf{v})) = \alpha \cdot f(h(\mathbf{v})) = \\ &= \alpha \cdot (f \circ h)(\mathbf{v}) = (\alpha \cdot (f \circ h))(\mathbf{v}) \end{aligned}$$

$$\begin{aligned} (f \circ (g + h))(\mathbf{v}) &= f((g + h)(\mathbf{v})) = f(g(\mathbf{v}) + h(\mathbf{v})) = \\ &= f(g(\mathbf{v})) + f(h(\mathbf{v})) = (f \circ g)(\mathbf{v}) + (f \circ h)(\mathbf{v}) = (f \circ g + f \circ h)(\mathbf{v}) \end{aligned}$$

¹ This terminology is not common, however, in this book we strictly follow this terminology.

$$\begin{aligned}
(f \circ (\alpha \cdot g))(\mathbf{v}) &= f((\alpha \cdot g)(\mathbf{v})) = f(\alpha \cdot g(\mathbf{v})) = \\
&= \alpha \cdot f(g(\mathbf{v})) = \alpha \cdot (f \circ g)(\mathbf{v}) = (\alpha \cdot (f \circ g))(\mathbf{v})
\end{aligned}$$

The above calculations prove the properties (9)-(12) of the composition of linear operators. $\square$

Let's fix the operator $h \in \text{End}(V)$ and consider the composition $f \circ h$ as a rule that maps each operator f to the other operator $g = f \circ h$. Then we get a mapping:

$$R_h : \text{End}(V) \rightarrow \text{End}(V).$$

The first two properties (9) and (10) from the theorem 1.1 mean that R_h is a linear mapping. This mapping is called the *right shift by h* since it acts as a composition, where h is placed on the right side. In a similar way we can define another mapping, which is called the *left shift by h* :

$$L_h : \text{End}(V) \rightarrow \text{End}(V).$$

It acts according to the rule $L_h(f) = h \circ f$. This mapping is linear due to the properties (11) and (12) from the theorem 1.1.

The operation of composition is an additional binary operation in the space of endomorphisms $\text{End}(V)$. The linearity of the mapping R_h is interpreted as the *linearity* of this binary operation in its *first argument*, while the *linearity* of L_h is said to be the linearity of composition in its *second argument*. A binary algebraic operation linear in both arguments is called a *bilinear* operation. A situation, where a linear vector space is equipped with an additional bilinear algebraic operation, is rather typical.

DEFINITION 1.1. A linear vector space A over a numeric field $\mathbb{K}$ equipped with a bilinear binary operation of vector multiplication is called an *algebra over the field $\mathbb{K}$* or simply a $\mathbb{K}$ -*algebra*.

The operation of multiplication in algebras is usually denoted by some sign like a dot $\langle \bullet \rangle$ or a circle $\langle \circ \rangle$, but very often this sign is omitted at all. The algebra A is called a *commutative algebra* if the multiplication in it is commutative: $ab = ba$. Similarly, the algebra A is called an *associative algebra* if the operation of multiplication is associative: $(ab)c = a(bc)$.

From the definition 1.1 and from the theorem 1.1 we conclude that the linear space $\text{End}(V)$ with the operation of composition taken for multiplication is an algebra over the same numeric field $\mathbb{K}$ as the initial vector space V . This algebra is called the *algebra of endomorphisms* of a linear vector space V . It is associative due to the theorem 1.6 from Chapter I. However, this algebra is not commutative in general case.

The operation of composition is treated as a multiplication in the algebra of endomorphisms $\text{End}(V)$. Therefore, it is usually omitted when written in this context. The multiplication of operator is higher priority operation as compared to addition. The priority of operator multiplication as compared to the multiplication by numbers makes no difference at all. This follows from the axiom (7) for the space $\text{End}(V)$ and from the properties (10) and (12) of the multiplication in $\text{End}(V)$. Now we can consider positive integer powers of linear operators:

$$f^2 = f f, \quad f^3 = f^2 f, \quad f^{n+1} = f^n f.$$

If an operator f is bijective, then we have the inverse operator f^{-1} and we can consider negative inverse powers of f as well:

$$f^{-2} = f^{-1} f^{-1}, \quad f^n f^{-n} = \text{id}_V, \quad f^{n+m} = f^n f^m.$$

The latter equality is valid either for positive and negative values of integer constants n and m .

DEFINITION 1.2. An algebra A over the field $\mathbb{K}$ is called an *algebra with unit element* or an *algebra with unity* if there is an element $1 \in A$ such that $1 \cdot a = a$ and $a \cdot 1 = a$ for all $a \in A$.

The algebra of endomorphisms $\text{End}(V)$ is an algebra with unity. The identical operator plays the role of unit element in this algebra: $1 = \text{id}_V$. Therefore, this operator is also called the *unit operator* or the *operator unity*.

DEFINITION 1.3. A linear operator $f: V \rightarrow V$ is called a *scalar operator* if it is obtained by multiplying the unit operator 1 by a number $\lambda \in \mathbb{K}$, i. e. if $f = \lambda \cdot 1$.

The basic purpose of operators from the space $\text{End}(V)$ is to act upon vectors of the space V . Suppose that $a, b \in \text{End}(V)$ and let $\mathbf{x}, \mathbf{y} \in V$. Then

- (1) $(a + b)(\mathbf{x}) = a(\mathbf{x}) + b(\mathbf{x})$;
- (2) $a(\mathbf{x} + \mathbf{y}) = a(\mathbf{x}) + a(\mathbf{y})$.

These two relationships are well known: the first one follows from the definition of the sum of two operators, the second relationship follows from the linearity of the operator a . The question is why the vectors $\mathbf{x}$ and $\mathbf{y}$ in the above formulas are surrounded by brackets. This is the consequence of «functional» form of writing the action of an operator upon a vector: the operator sign is put on the left and the vector sign is put on the right and is enclosed into brackets like an argument of a function: $\mathbf{w} = f(\mathbf{v})$. Algebraists use the more «deliberate» form of writing: $\mathbf{w} = f \mathbf{v}$. The operator sign is on the left and the vector sign on the right, but no brackets are used. If we know that $f \in \text{End}(V)$ and $\mathbf{v} \in V$, then such a writing makes no confusion. In more complicated case even if we know that $\alpha \in \mathbb{K}$, $f, g \in \text{End}(V)$, and $\mathbf{v} \in V$, the writing $\mathbf{w} = \alpha f g \mathbf{v}$ admits several interpretations:

$$\begin{aligned} \mathbf{w} &= \alpha \cdot f(g(\mathbf{v})), & \mathbf{w} &= (\alpha \cdot f)(g(\mathbf{v})), \\ \mathbf{w} &= (\alpha \cdot (f \circ g))(\mathbf{v}), & \mathbf{w} &= ((\alpha \cdot f) \circ g)(\mathbf{v}). \end{aligned}$$

However, for any one of these interpretations we get the same vector $\mathbf{w}$. Therefore, in what follows we shall use the algebraic form of writing the action of an operator upon a vector, especially in huge calculations.

Let $f: V \rightarrow V$ be a linear operator in a finite-dimensional vector space V . According to general scheme of constructing the matrix of a linear mapping we should choose two bases $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $\mathbf{h}_1, \dots, \mathbf{h}_n$ in V and consider the expansions similar to (9.1) in Chapter I. No doubt that this approach is valid, it could be very fruitful in some cases. However, to have two bases in one space — it is certainly excessive. Therefore, when constructing the matrix of a linear operator the second basis $\mathbf{h}_1, \dots, \mathbf{h}_n$ is chosen to be coinciding with the first one. The

$$\begin{aligned} f(\mathbf{e}_1) &= F_1^1 \cdot \mathbf{e}_1 + \dots + F_1^n \cdot \mathbf{e}_n, \\ \vdots \\ f(\mathbf{e}_n) &= F_n^1 \cdot \mathbf{e}_1 + \dots + F_n^n \cdot \mathbf{e}_n, \end{aligned} \tag{1.1}$$
$$f(\mathbf{e}_j) = \sum_{i=1}^n F_j^i \cdot \mathbf{e}_i. \quad (1.2)$$

- (1) the sum of two operators is represented by the sum of their matrices;
- (2) the product of an operator by a number is represented by the product of its matrix by that number;
- (3) the composition of two matrices is represented by the product of their matrices.

$$\begin{aligned} h(\mathbf{e}_j) &= (f + g) \mathbf{e}_j = f(\mathbf{e}_j) + h(\mathbf{e}_j) = \\ &= \sum_{i=1}^n F_j^i \cdot \mathbf{e}_i + \sum_{i=1}^n G_j^i \cdot \mathbf{e}_i = \sum_{i=1}^n (F_j^i + G_j^i) \cdot \mathbf{e}_i = \sum_{i=1}^n H_j^i \cdot \mathbf{e}_i. \end{aligned}$$
$$\begin{aligned} h(\mathbf{e}_j) &= (\alpha \cdot f) \mathbf{e}_j = \alpha \cdot f(\mathbf{e}_j) = \\ &= \alpha \cdot \left(\sum_{i=1}^n F_j^i \cdot \mathbf{e}_i \right) = \sum_{i=1}^n (\alpha F_j^i) \cdot \mathbf{e}_i = \sum_{i=1}^n H_j^i \cdot \mathbf{e}_i. \end{aligned}$$
$$\begin{aligned} h(\mathbf{e}_j) &= (f \circ g) \mathbf{e}_j = f(g(\mathbf{e}_g)) = f\left(\sum_{i=1}^n G_j^i \cdot \mathbf{e}_i\right) = \sum_{i=1}^n G_j^i \cdot f(\mathbf{e}_i) = \\ &= \sum_{i=1}^n G_j^i \cdot \left(\sum_{s=1}^n F_i^s \cdot \mathbf{e}_s\right) = \sum_{s=1}^n \left(\sum_{i=1}^n F_i^s G_j^i\right) \cdot \mathbf{e}_s = \sum_{s=1}^n H_i^s \cdot \mathbf{e}_s. \end{aligned}$$

Due to the uniqueness of the expansion of a vector in a basis we derive

$$H_i^s = \sum_{j=1}^n F_i^s G_j^i.$$

The right side of this equality is easily interpreted as the product of two matrices written in terms of the components of these matrices. Therefore, $H = FG$. The theorem is proved. $\square$

From the theorem that was proved just above we conclude that when relating an operator $f \in \text{End}(V)$ with its matrix we establish the isomorphism of the algebra $\text{End}(V)$ and the matrix algebra $\mathbb{K}^{n \times n}$ with standard matrix multiplication.

Now let's study how the matrix of a linear operator $f: V \rightarrow V$ changes under the change of the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ for some other basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$. Let S be the direct transition matrix and let T be the inverse one. Note that we need not derive the transformation formulas again. We can adapt the formulas (9.10) from Chapter I for our present purpose. Since the basis $\mathbf{h}_1, \dots, \mathbf{h}_n$ coincides with $\mathbf{e}_1, \dots, \mathbf{e}_n$ and the basis $\tilde{\mathbf{h}}_1, \dots, \tilde{\mathbf{h}}_n$ coincides with $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$, we have $P = S$. Then transformation formulas are written as

$$\tilde{F} = S^{-1} F S, \quad F = S \tilde{F} S^{-1}. \quad (1.3)$$

These are the required formulas for transforming the matrix of a linear operator under a change of basis. Taking into account that $T = S^{-1}$ we can write (1.3) as

$$\tilde{F}_p^q = \sum_{i=1}^n \sum_{j=1}^n T_i^q S_p^j F_j^i, \quad F_j^i = \sum_{q=1}^n \sum_{p=1}^n S_q^i T_j^p F_p^q. \quad (1.4)$$

The relationships (1.3) yield very important formula relating the determinants of the matrices F and $\tilde{F}$. Indeed, we have

$$\det \tilde{F} = \det(S^{-1}) \det F \det S = (\det S)^{-1} \det F \det S = \det F.$$

The coincidence of determinants of the matrices of a linear operator f in two arbitrary bases mean that they represent a number which does not depend on a basis at all.

DEFINITION 1.4. The *determinant* $\det f$ of a linear operator f is the number equal to the determinant of the matrix F of this linear operator in some basis.

A *numeric invariant* of a geometric object in a linear vector space V is a number determined by this geometric object such that it does not depend on anything else other than that geometric object itself. The determinant of a linear operator $\det f$ is an example of such numeric invariant. Coordinates of a vector or components of the matrix of a linear operator are not numeric invariants. Another example of a numeric invariant of a linear operator is its rank:

$$\text{rank } f = \dim(\text{Im } f).$$

Soon we shall define a lot of other numeric invariants of a linear operator.

From the third proposition of the theorem 1.2 we derive the following formula for the determinant of a linear operator:

$$\det(f \circ g) = \det(f) \cdot \det(g). \quad (1.5)$$

THEOREM 1.3. *A linear operator $f : V \rightarrow V$ in a finite-dimensional linear vector space V is injective if and only if it is surjective.*

PROOF. In order to prove this theorem we apply the theorem 1.2 and two theorems 8.3 and 9.4 from Chapter I. The injectivity of the linear operator f is equivalent to the condition $\text{Ker } f = \{\mathbf{0}\}$, the surjectivity of the operator f is equivalent to $\text{Im } f = V$, while the theorem 9.4 from Chapter I relates the dimensions of these two subspaces $\text{Ker } f$ and $\text{Im } f$:

$$\dim(\text{Ker } f) + \dim(\text{Im } f) = \dim(V).$$

If the operator f is injective, then $\text{Ker } f = \{\mathbf{0}\}$ and $\dim(\text{Ker } f) = 0$. Then $\dim(\text{Im } f) = \dim(V)$. Applying the third proposition of the theorem 4.5 from Chapter I, we get $\text{Im } f = V$, which proves the surjectivity of the operator f .

Conversely, if the operator f is surjective, then $\text{Im } f = V$ and $\dim(\text{Im } f) = \dim(V)$. Hence, $\dim(\text{Ker } f) = 0$ and $\text{Ker } f = \{\mathbf{0}\}$. This proves the injectivity of the operator f . $\square$

THEOREM 1.4. *A linear operator $f : V \rightarrow V$ in a finite-dimensional linear vector space V is bijective if and only if $\det f \neq 0$.*

PROOF. Let $\mathbf{x}$ be a vector of V and let $\mathbf{y} = f(\mathbf{x})$. Expanding $\mathbf{x}$ and $\mathbf{y}$ in some basis $\mathbf{e}_1, \dots, \mathbf{e}_n$, we get the following formula relating their coordinates:

$$\left\| \begin{matrix} y^1 \\ \vdots \\ y^n \end{matrix} \right\| = \left\| \begin{matrix} F_1^1 & \dots & F_n^1 \\ \vdots & \ddots & \vdots \\ F_1^n & \dots & F_n^n \end{matrix} \right\| \cdot \left\| \begin{matrix} x^1 \\ \vdots \\ x^n \end{matrix} \right\|. \quad (1.6)$$

The formula (1.6) can be derived independently or one can derive it from the formula (9.5) of Chapter I. From this formula we derive that $\mathbf{x}$ belong to the kernel of the operator f if and only if its coordinates $x^1, \dots, x^n$ satisfy the homogeneous system of linear equations

$$\left\| \begin{matrix} F_1^1 & \dots & F_n^1 \\ \vdots & \ddots & \vdots \\ F_1^n & \dots & F_n^n \end{matrix} \right\| \cdot \left\| \begin{matrix} x^1 \\ \vdots \\ x^n \end{matrix} \right\| = \left\| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right\|, \quad (1.7)$$

The matrix of this system of equations coincides with the matrix of the operator f in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. Therefore, the kernel of the operator f is nonzero if and only if the system of equations (1.7) has nonzero solution. Here we use the well-known result from the theory of determinants: a homogeneous system of linear algebraic equations with square matrix F has nonzero solution if and only if $\det F = 0$. The proof of this fact can be found in [5]. From this result immediately get that the condition $\text{Ker } f \neq \{\mathbf{0}\}$ is equivalent to $\text{Ker } f = \{\mathbf{0}\}$. Due to the previous theorem and due to the theorem 1.1 from Chapter I the latter equality $\text{Ker } f \neq \{\mathbf{0}\}$ is equivalent to bijectivity of f . The theorem is proved. $\square$

An operator f with zero determinant $\det f = 0$ is called a *degenerate operator*. Using this terminology we can formulate the following corollary of the theorem 1.4.

COROLLARY. *A linear operator $f: V \rightarrow V$ in a finite-dimensional space V has a nontrivial kernel $\text{Ker } f \neq \{0\}$ if and only if it is degenerate. Otherwise this linear operator is bijective.*

Remember that a bijective linear mapping f from V to W is called an isomorphism. If $W = V$ such a mapping establishes an isomorphism of the space V with itself. Therefore, it is called an *automorphism* of the space V . The set of all automorphisms of the space V is denoted by $\text{Aut}(V)$. It is obvious that $\text{Aut}(V)$ possesses the following properties:

- (1) if $f, g \in \text{Aut}(V)$, then $f \circ g \in \text{Aut}(V)$;
- (2) if $f \in \text{Aut}(V)$, then $f^{-1} \in \text{Aut}(V)$;
- (3) $1 \in \text{Aut}(V)$, where 1 is the identical operator.

It is easy to see that due to the above three properties the set of automorphisms $\text{Aut}(V)$ is equipped with a structure of a group. *The group of automorphisms* $\text{Aut}(V)$ is a subset in the algebra of endomorphisms $\text{End}(V)$, however, it does not inherit the structure of an algebra, nor even the structure of a linear vector space. It is clear because, for instance, the zero operator does not belong to $\text{Aut}(V)$. In the case of finite-dimensional space V the group of automorphisms consists of all non-degenerate operators.

§ 2. Projection operators.

Let V be a linear vector space expanded into a direct sum of two subspaces:

$$V = U_1 \oplus U_2. \quad (2.1)$$

Due to the expansion (2.1) each vector $\mathbf{v} \in V$ is expanded into a sum

$$\mathbf{v} = \mathbf{u}_1 + \mathbf{u}_2, \text{ where } \mathbf{u}_1 \in U_1 \text{ and } \mathbf{u}_2 \in U_2, \quad (2.2)$$

the components $\mathbf{u}_1$ and $\mathbf{u}_2$ in (2.2) being uniquely determined by the vector $\mathbf{v}$.

DEFINITION 2.1. The operator $P: V \rightarrow V$ mapping each vector $v \in V$ to its first component $\mathbf{u}_1$ in the expansion (2.2) is called the *operator of projection* onto the subspace U_1 parallel to the subspace U_2 .

THEOREM 2.1. *For any expansion of the form (2.1) the operator of projection onto the subspace U_1 parallel to the subspace U_2 is a linear operator.*

PROOF. Let's consider a pair of vectors $\mathbf{v}_1, \mathbf{v}_2$ from the space V , and for each of them consider the expansion like (2.2):

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{u}_1 + \mathbf{u}_2, \\ \mathbf{v}_2 &= \tilde{\mathbf{u}}_1 + \tilde{\mathbf{u}}_2. \end{aligned}$$

Then $P(\mathbf{v}_1) = \mathbf{u}_1$ and $P(\mathbf{v}_2) = \tilde{\mathbf{u}}_1$. Let's add the above two expansions and write

$$\mathbf{v}_1 + \mathbf{v}_2 = (\mathbf{u}_1 + \tilde{\mathbf{u}}_1) + (\mathbf{u}_2 + \tilde{\mathbf{u}}_2). \quad (2.3)$$

From $\mathbf{u}_1, \tilde{\mathbf{u}}_1 \in U_1$ and from $\mathbf{u}_2, \tilde{\mathbf{u}}_2 \in U_2$ we derive $\mathbf{u}_1 + \tilde{\mathbf{u}}_1 \in U_1$ and $\mathbf{u}_2 + \tilde{\mathbf{u}}_2 \in U_2$. Therefore, (2.3) is an expansion of the form (2.2) for the vector $\mathbf{v}_1 + \mathbf{v}_2$. Then

$$P(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{u}_1 + \tilde{\mathbf{u}}_1 = P(\mathbf{v}_1) + P(\mathbf{v}_2). \quad (2.4)$$

Now let's consider the expansion (2.2) for an arbitrary vector $\mathbf{v} \in V$ and multiply it by a number $\alpha \in \mathbb{K}$:

$$\alpha \cdot \mathbf{v} = (\alpha \cdot \mathbf{u}_1) + (\alpha \cdot \mathbf{u}_2).$$

Then $\alpha \cdot \mathbf{u}_1 \in U_1$ and $\alpha \cdot \mathbf{u}_2 \in U_2$, therefore, due to the definition of P we get

$$P(\alpha \cdot \mathbf{v}) = \alpha \cdot \mathbf{u}_1 = \alpha \cdot P(\mathbf{v}). \quad (2.5)$$

The relationships (2.4) and (2.5) are just the very relationships that mean the linearity of the operator P . $\square$

Suppose that $\mathbf{v}$ in the expansion (2.2) is chosen to be a vector of the subspace U_1 . Then the expansion (2.2) for this vector is $\mathbf{v} = \mathbf{v} + \mathbf{0}$, therefore, $P(\mathbf{v}) = \mathbf{v}$. This means that all vectors of the subspace U_1 are projected by P onto themselves. This fact has an important consequence $P^2 = P$. Indeed, for any $\mathbf{v} \in V$ we have $P(\mathbf{v}) \in U_1$, therefore, $P(P(\mathbf{v})) = P(\mathbf{v})$.

Besides P , by means of (2.2) we can define the other operator Q such that $Q(\mathbf{v}) = \mathbf{u}_2$. It is also a projection operator: it projects onto U_2 parallel to U_1 . Therefore, $Q^2 = Q$. For the sum of these two operators we get $P + Q = 1$. Indeed, for any vector $v \in V$ we have

$$P(\mathbf{v}) + Q(\mathbf{v}) = \mathbf{u}_1 + \mathbf{u}_2 = \mathbf{v} = \text{id}_V(\mathbf{v}) = 1(\mathbf{v}).$$

If $\mathbf{v} \in U_1$, then the expansion (2.2) for this vector is $\mathbf{v} = \mathbf{v} + \mathbf{0}$, therefore, $Q(\mathbf{v}) = \mathbf{0}$. Similarly, $P(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in U_2$. Hence, we derive $Q(P(\mathbf{v})) = \mathbf{0}$ and $P(Q(\mathbf{v})) = \mathbf{0}$ for any $v \in V$. Summarizing these results, we write

$$\begin{aligned} P^2 &= P, & P + Q &= 1, \\ Q^2 &= Q, & PQ &= QP = 0. \end{aligned} \quad (2.6)$$

A pair of projection operators satisfying the relationships (2.6) is called a *concordant pair of projectors*.

In order to get a concordant pair of projectors it is sufficient to define only one of them, for instance, the operator P . The second operator Q then is given by formula $Q = 1 - P$. All of the relationships (2.6) thereby will be automatically fulfilled. Indeed, we have the relationships

$$\begin{aligned} PQ &= P \circ (1 - P) = P - P^2 = P - P = 0, \\ QP &= (1 - P) \circ P = P - P^2 = P - P = 0. \end{aligned}$$

The relationship $Q^2 = Q$ for Q is derived in a similar way:

$$Q^2 = (1 - P) \circ (1 - P) = 1 - 2P + P = 1 - P = Q.$$

THEOREM 2.2. *An operator $P: V \rightarrow V$ is a projector onto a subspace parallel to another subspace if and only if $P^2 = P$.*

PROOF. We have already shown that any projector satisfies the equality $P^2 = P$. Let's prove the converse proposition. Suppose that $P^2 = P$. Let's denote $Q = 1 - P$. Then for operators P and Q all of the relationships (2.6) are fulfilled. Let's consider two subspaces

$$U_1 = \text{Im } P, \quad U_2 = \text{Ker } P.$$

For an arbitrary vector $\mathbf{v} \in V$ we have the expansion

$$\mathbf{v} = 1(\mathbf{v}) = (P + Q)\mathbf{v} = P(\mathbf{v}) + Q(\mathbf{v}), \quad (2.7)$$

where $\mathbf{u}_1 = P(\mathbf{v}) \in \text{Im } P$. From the relationship $PQ = 0$ for the other vector $\mathbf{u}_2 = Q(\mathbf{v})$ in (2.7) we get the equality

$$P(\mathbf{u}_2) = P(Q(\mathbf{v})) = \mathbf{0}.$$

This means $\mathbf{u}_2 \in \text{Ker } P$. Hence, $V = \text{Im } P + \text{Ker } P$. Let's prove that this is a direct sum of subspaces. We should prove the uniqueness of the expansion

$$\mathbf{v} = \mathbf{u}_1 + \mathbf{u}_2, \quad (2.8)$$

where $\mathbf{u}_1 \in \text{Im } P$ and $\mathbf{u}_2 \in \text{Ker } P$. From $\mathbf{u}_1 \in \text{Im } P$ we conclude that $\mathbf{u}_1 = P(\mathbf{v}_1)$ for some vector $\mathbf{v}_1 \in V$. From $\mathbf{u}_2 \in \text{Ker } P$ we derive $P(\mathbf{u}_2) = \mathbf{0}$. Then from (2.8) we derive the following formulas:

$$\begin{aligned} P(\mathbf{v}) &= P(\mathbf{u}_1) + P(\mathbf{u}_2) = P(P(\mathbf{v}_1)) = P^2(\mathbf{v}_1) = P(\mathbf{v}_1) = \mathbf{u}_1, \\ Q(\mathbf{v}) &= (1 - P)\mathbf{v} = \mathbf{v} - P(\mathbf{v}) = \mathbf{v} - \mathbf{u}_1 = \mathbf{u}_2. \end{aligned}$$

The relationships derived just above mean that any expansion (2.8) coincides with (2.7). Hence, it is unique and we have

$$V = \text{Im } P \oplus \text{Ker } P.$$

The operator P maps an arbitrary vector $v \in V$ into the first component of the expansion (2.8). Hence, P is an operator of projection onto the subspace $\text{Im } P$ parallel to the subspace $\text{Ker } P$. $\square$

Now suppose that a linear vector space V is expanded into the direct sum of several its subspaces $U_1, \dots, U_s$:

$$V = U_1 \oplus \dots \oplus U_s. \quad (2.9)$$

This expansion of the space V implies the unique expansion for each vector $\mathbf{v} \in V$:

$$\mathbf{v} = \mathbf{u}_1 + \dots + \mathbf{u}_s, \quad \text{where } \mathbf{u}_i \in U_i \quad (2.10)$$

DEFINITION 2.2. The operator $P_i: V \rightarrow V$ that maps each vector $\mathbf{v} \in V$ to its i -th component $\mathbf{u}_i$ in the expansion (2.10) is called the *operator of projection* onto U_i parallel to other subspaces.

The proof of linearity of the operators P_i is practically the same as in case of two subspaces considered in theorem 2.1. It is based on the uniqueness of the expansion (2.10).

Let's choose a vector $\mathbf{u} \in U_i$. Then its expansion (2.10) looks like:

$$\mathbf{u} = \mathbf{0} + \dots + \mathbf{0} + \mathbf{u} + \mathbf{0} + \dots + \mathbf{0}.$$

Therefore for any such vector $\mathbf{u}$ we have $P_i(\mathbf{u}) = \mathbf{u}$ and $P_j(\mathbf{u}) = \mathbf{0}$ for $j \neq i$. For the projection operators P_i this yields

$$(P_i)^2 = P_i, \quad P_i \circ P_j = 0 \quad \text{for } i \neq j. \quad (2.11)$$

Moreover, from the definition of P_i we get

$$P_1 + \dots + P_s = 1. \quad (2.12)$$

Due to the first relationship (2.11) the theory of separate operators P_i does not differ from the theory of projectors defined by two component expansions of the space V . In the case of multicomponent expansions the collective behavior of projectors is of particular interest. A family of projection operators $P_1, \dots, P_s$ is called a *concordant family of projectors* if the operators of this family satisfy the relationships (2.11) and (2.12).

THEOREM 2.3. A family of projection operators $P_1, \dots, P_s$ is determined by an expansion of the form (2.9) if and only if it is concordant, i. e. if these operators satisfy the relationships (2.11) and (2.12).

PROOF. We already know that a family of projectors determined by an expansion (2.9) satisfy the relationships (2.11) and (2.12). Let's prove the converse proposition. Suppose that we have a family of operators $P_1, \dots, P_s$ satisfying the relationships (2.11) and (2.12). Then we define the subspaces $U_i = \text{Im } P_i$. Due to the relationship (2.12) for an arbitrary vector $\mathbf{v} \in V$ we get

$$\mathbf{v} = P_1(\mathbf{v}) + \dots + P_s(\mathbf{v}), \quad (2.13)$$

where $P_i(\mathbf{v}) \in \text{Im } P_i$. Hence, we have the expansion of V into a sum of subspaces

$$V = \text{Im } P_1 + \dots + \text{Im } P_s. \quad (2.14)$$

Let's prove that the sum (2.14) is a direct sum. For this purpose we consider an expansion of some arbitrary vector $\mathbf{v} \in V$ corresponding to the expansion (2.14):

$$\mathbf{v} = \mathbf{u}_1 + \dots + \mathbf{u}_s, \quad \text{where } \mathbf{u}_i \in \text{Im } P_i \quad (2.15)$$

From $\mathbf{u}_i \in \text{Im } P_i$ we conclude that $\mathbf{u}_i = P(\mathbf{v}_i)$, where $\mathbf{v}_i \in V$. Then from the expansion (2.15) we derive the following equality:

$$P_i(\mathbf{v}) = P_i(\mathbf{u}_1 + \dots + \mathbf{u}_s) = \sum_{j=1}^s P_i(P_j(\mathbf{v}_j)).$$

Due to (2.11) only one term in the above sum is nonzero. Therefore, we have

$$P_i(\mathbf{v}) = (P_i)^2 \mathbf{v}_i = P_i(\mathbf{v}_i) = \mathbf{u}_i.$$

This equality shows that an arbitrary expansion (2.15) should coincide with (2.13). This means that (2.13) is the unique expansion of that sort. Hence, the sum (2.14) is a direct sum and P_i is the projection operator onto the i -th component of the sum (2.14) parallel to its other components. The theorem is proved. $\square$

Now we consider a projection operator P as an example for the first approach to the problem of bringing the matrix of a linear operator to a canonic form.

THEOREM 2.4. *For any nonzero projection operator in a finite-dimensional vector space V there is a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ such that the matrix of the operator P has the following form in that basis:*

$$\mathcal{P} = \left\| \begin{array}{cccccc} \overbrace{\begin{matrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \end{matrix}}^s & & & & & \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{array} \right\|_s. \quad (2.16)$$

PROOF. Let's consider the subspaces $\text{Im } P$ and $\text{Ker } P$. From the condition $P \neq 0$ we conclude that $s = \dim(\text{Im } P) \neq 0$. Then we choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ in $U_1 = \text{Im } P$ and if $U_1 \neq V$, we complete it by choosing a basis in $\mathbf{e}_{s+1}, \dots, \mathbf{e}_n$ in $U_2 = \text{Ker } P$. The sum of these two subspaces is a direct sum: $V = U_1 \oplus U_2$, therefore, joining together two bases in them, we get a basis of V (see the proof of theorem 6.3 in Chapter I).

Now let's apply the operator P to the vectors of the basis we have constructed just above. This operator projects onto U_1 parallel to U_2 , therefore, we have

$$P(\mathbf{e}_i) = \begin{cases} \mathbf{e}_i & \text{for } i = 1, \dots, s, \\ \mathbf{0} & \text{for } i = s+1, \dots, n. \end{cases}$$

Due to this formula it's clear that (2.16) is the matrix of the projection operator P the basis $\mathbf{e}_1, \dots, \mathbf{e}_s$. $\square$

§ 3. Invariant subspaces.
Restriction and factorization of operators.

Let $f: V \rightarrow V$ be a linear operator and let U be a subspace of V . Let's restrict the domain of f to the subspace U . Thereby the image of f shrinks to $f(U)$. However, in general, the subspace $f(U)$ is not enclosed into the subspace U . For this reason in general case we should treat the *restricted* operator f as a linear mapping $f_U: U \rightarrow V$, rather than a linear operator.

DEFINITION 3.1. A subspace U is called an *invariant subspace* of a linear operator $f: V \rightarrow V$ if $f(U) \subseteq U$, i.e. if $\mathbf{u} \in U$ implies $f(\mathbf{u}) \in U$.

If U is an invariant subspace of f , the restriction f_U can be treated as a linear operator in U . Its action upon vectors $\mathbf{u} \in U$ coincides with the action of f upon $\mathbf{u}$. As for the vectors outside the subspace U , the operator f_U cannot be applied to them at all.

THEOREM 3.1. *The kernel and the image of a linear operator $f: V \rightarrow V$ are invariant subspaces of f .*

PROOF. Let's consider the kernel of f for the first. If $\mathbf{u} \in \text{Ker } f$, then $f(\mathbf{u}) = \mathbf{0}$. Hence, $f(\mathbf{u}) \in U$, since the zero vector $\mathbf{0}$ is an element of any subspace of V . The invariance of the kernel $\text{Ker } f$ is proved.

Now let $\mathbf{u} \in \text{Im } f$. Denote $\mathbf{w} = f(\mathbf{u})$. Then $\mathbf{w}$ is the image of the vector $\mathbf{u}$, hence, $\mathbf{w} = f(\mathbf{u}) \in \text{Im } f$. The invariance of the image $\text{Im } f$ is proved. $\square$

THEOREM 3.2. *The intersection and the sum of an arbitrary number of invariant subspaces of a linear operator $f: V \rightarrow V$ both are the invariant subspaces of f .*

PROOF. Let U_i , $i \in I$ be a family of invariant subspaces of a linear operator $f: V \rightarrow V$. Let's consider the intersection and the sum of these subspaces:

$$U = \bigcap_{i \in I} U_i, \quad W = \sum_{i \in I} U_i.$$

In § 6 of Chapter I we have proved that U and W are the subspaces of V . Now we should prove that they are invariant subspaces. For the first, let's prove that U is an invariant subspace. Consider a vector $\mathbf{u} \in U$. This vector belongs to all subspaces U_i , which are invariant subspaces of f . Therefore, $f(\mathbf{u})$ also belongs to all subspaces U_i . This means that $f(\mathbf{u})$ belongs to their intersection U . The invariance of U is proved.

Now let's consider a vector $\mathbf{w} \in W$. According to the definition of the sum of subspaces, this vector admits the expansion

$$\mathbf{w} = \mathbf{u}_{i_1} + \dots + \mathbf{u}_{i_s}, \text{ where } \mathbf{u}_{i_r} \in U_{i_r}.$$

Applying the operator f to both sides of this equality, we get:

$$f(\mathbf{w}) = f(\mathbf{u}_{i_1}) + \dots + f(\mathbf{u}_{i_s}).$$

Due to the invariance of U_i we have $f(\mathbf{u}_{i_r}) \in U_{i_r}$. Hence, $f(\mathbf{w}) \in W$. This yields the invariance of the sum W of the invariant subspaces U_i . $\square$

Let U be an invariant subspace of a linear operator $f: V \rightarrow V$. Let's consider the factorspace V/U and define the operator $f_{V/U}$ in this factorspace by formula

$$f_{V/U}(Q) = \text{Cl}_U(f(\mathbf{v})), \text{ where } Q = \text{Cl}_U(\mathbf{v}). \quad (3.1)$$

The operator $f_{V/U}: V/U \rightarrow V/U$ acting according to the rule (3.1) is called the *factoroperator* of the *quotient operator* of the operator f by the subspace U . We can rewrite the formula (3.1) in shorter form as follows:

$$f_{V/U}(\text{Cl}_U(\mathbf{v})) = \text{Cl}_U(f(\mathbf{v})). \quad (3.2)$$

Like formulas (7.3) in Chapter I, the formulas (3.1) and (3.2) comprise the definite amount of uncertainty due to the uncertainty of the choice of a representative $\mathbf{v}$ in a coset $Q = \text{Cl}_U(\mathbf{v})$. Therefore, we need to prove their correctness.

THEOREM 3.3. *The formula (3.1) and the equivalent formula (3.2) both are correct. They define a linear operator $f_{V/U}$ in factorspace V/U .*

PROOF. Let's consider two different representative vectors in a coset Q , i.e. let $\mathbf{v}, \tilde{\mathbf{v}} \in Q$. Then $\tilde{\mathbf{v}} - \mathbf{v} \in U$. According to the formula (3.1), we consider two possible results of applying the operator $f_{V/U}$ to Q :

$$f_{V/U}(Q) = \text{Cl}_U(f(\mathbf{v})), \quad f_{V/U}(Q) = \text{Cl}_U(f(\tilde{\mathbf{v}})).$$

Let's calculate the difference of these two possible results:

$$\text{Cl}_U(f(\tilde{\mathbf{v}})) - \text{Cl}_U(f(\mathbf{v})) = \text{Cl}_U(f(\tilde{\mathbf{v}}) - f(\mathbf{v})) = \text{Cl}_U(f(\tilde{\mathbf{v}} - \mathbf{v})).$$

Note that the vector $\mathbf{u} = \tilde{\mathbf{v}} - \mathbf{v}$ belongs to the subspace U . Since U is an invariant subspace, we have $\tilde{\mathbf{u}} = f(\mathbf{u}) \in U$. Therefore, we get

$$\text{Cl}_U(f(\tilde{\mathbf{v}})) - \text{Cl}_U(f(\mathbf{v})) = \text{Cl}_U(\tilde{\mathbf{u}}) = \mathbf{0}.$$

This coincidence $\text{Cl}_U(f(\tilde{\mathbf{v}})) = \text{Cl}_U(f(\mathbf{v}))$ that we have proved just above proves the correctness of the formula (3.1) and the formula (3.2) as well.

Now let's prove the linearity of the factoroperator $f_{V/U}: V/U \rightarrow V/U$. We shall carry out the appropriate calculations on the base of formula (3.1):

$$\begin{aligned} f_{V/U}(Q_1 + Q_2) &= f_{V/U}(\text{Cl}_U(\mathbf{v}_1) + \text{Cl}_U(\mathbf{v}_2)) = \\ &= f_{V/U}(\text{Cl}_U(\mathbf{v}_1 + \mathbf{v}_2)) = \text{Cl}_U(f(\mathbf{v}_1 + \mathbf{v}_2)) = \\ &= \text{Cl}_U(f(\mathbf{v}_1)) + \text{Cl}_U(f(\mathbf{v}_2)) = f_{V/U}(Q_1) + f_{V/U}(Q_2), \end{aligned}$$

$$\begin{aligned} f_{V/U}(\alpha \cdot Q) &= f_{V/U}(\alpha \cdot \text{Cl}_U(\mathbf{v})) = \\ &= f_{V/U}(\text{Cl}_U(\alpha \cdot \mathbf{v})) = \text{Cl}_U(f(\alpha \cdot \mathbf{v})) = \\ &= \text{Cl}_U(\alpha \cdot f(\mathbf{v})) = \alpha \cdot \text{Cl}_U(f(\mathbf{v})) = \alpha \cdot f_{V/U}(Q). \end{aligned}$$

These calculations show that $f_{V/U}$ is a linear operator. The theorem is proved. $\square$

THEOREM 3.4. *Suppose that U is a common invariant subspace of two linear operators $f, g \in \text{End}(V)$. Then U is an invariant subspace of the operators $f+g$, $\alpha \cdot f$ and $f \circ g$ as well. For their restrictions to the subspace U and for the corresponding factoroperators we have the following relationships:*

$$\begin{aligned} (f+g)_U &= f_U + g_U; & (f+g)_{V/U} &= f_{V/U} + g_{V/U}; \\ (\alpha \cdot f)_U &= \alpha \cdot f_U; & (\alpha \cdot f)_{V/U} &= \alpha \cdot f_{V/U}; \\ (f \circ g)_U &= f_U \circ g_U; & (f \circ g)_{V/U} &= f_{V/U} \circ g_{V/U}. \end{aligned}$$

PROOF. Let's begin with the first case. Denote $h = f + g$ and assume that $\mathbf{u}$ is an arbitrary vector of U . Then $f(\mathbf{u}) \in U$ and $g(\mathbf{u}) \in U$ since U is an invariant subspace of both operators f and g . For this reason we obtain $h(\mathbf{u}) = f(\mathbf{u}) + g(\mathbf{u}) \in U$. This proves that U is an invariant subspace of h . The relationship $h_U = f_U + g_U$ follows from $h = f + g$ since the results of applying the restricted operators to $\mathbf{u}$ do not differ from the results of applying f , g , and h to $\mathbf{u}$. The corresponding relationship for the factoroperators is proved as follows:

$$\begin{aligned} h_{V/U}(\text{Cl}_U(\mathbf{v})) &= \text{Cl}_U(h(\mathbf{v})) = \text{Cl}_U(f(\mathbf{v}) + h(\mathbf{v})) = \\ &= \text{Cl}_U(f(\mathbf{v})) + \text{Cl}_U(g(\mathbf{v})) = f_{V/U}(\text{Cl}_U(\mathbf{v})) + g_{V/U}(\text{Cl}_U(\mathbf{v})). \end{aligned}$$

The second case, where we denote $h = \alpha \cdot f$, is not quite different from the first one. From $\mathbf{u} \in U$ it follows that $f(\mathbf{u}) \in U$, hence, $h(\mathbf{u}) = \alpha \cdot f(\mathbf{u}) \in U$. The relationship $h_U = \alpha \cdot f_U$ now is obvious due to the same reasons as above. For the factoroperators we perform the following calculations:

$$\begin{aligned} h_{V/U}(\text{Cl}_U(\mathbf{v})) &= \text{Cl}_U(h(\mathbf{v})) = \text{Cl}_U(\alpha \cdot f(\mathbf{v})) = \\ &= \alpha \cdot \text{Cl}_U(f(\mathbf{v})) = \alpha \cdot f_{V/U}(\text{Cl}_U(\mathbf{v})) = (\alpha \cdot f_{V/U})(\text{Cl}_U(\mathbf{v})). \end{aligned}$$

Now we consider the third case. Here we denote $h = f \circ g$. From $\mathbf{u} \in U$ we derive $\mathbf{w} = g(\mathbf{u}) \in U$, then from $\mathbf{w} \in U$ we derive $f(\mathbf{w}) \in U$, which means that U is an invariant subspace of h . Indeed, $h(\mathbf{u}) = f(g(\mathbf{u})) = f(\mathbf{w}) \in U$. For the restricted operators this yields the equality

$$h_U(\mathbf{u}) = h(\mathbf{u}) = f(g(\mathbf{u})) = f_U(g_U(\mathbf{u})).$$

Hence, $h_U = f_U \circ g_U$. Passing to factoroperators, we obtain

$$\begin{aligned} h_{V/U}(\text{Cl}_U(v)) &= \text{Cl}_U(h(\mathbf{v})) = \text{Cl}_U(f(g(\mathbf{v}))) = f_{V/U}(\text{Cl}_U(g(\mathbf{v}))) = \\ &= f_{V/U}(g_{V/U}(\text{Cl}_U(\mathbf{v}))) = f_{V/U} \circ g_{V/U}(\text{Cl}_U(\mathbf{v})). \end{aligned}$$

The above calculations prove the last relationship of the theorem 3.4. $\square$

THEOREM 3.5. *Let $V = U_1 \oplus \dots \oplus U_s$ be an expansion of a linear vector space V into a direct sum of its subspaces. The subspaces $U_1, \dots, U_s$ are invariant sub-*

spaces of an operator $f: V \rightarrow V$ if and only if the projection operators $P_1, \dots, P_s$ associated with the expansion $V = U_1 \oplus \dots \oplus U_s$ commute with the operator f , i. e. if $f \circ P_i = P_i \circ f$, where $i = 1, \dots, s$.

PROOF. Suppose that all subspaces U_i are invariant under the action of the operator f . For an arbitrary vector $\mathbf{v} \in V$ we consider the expansion determined by the direct sum $V = U_1 \oplus \dots \oplus U_s$:

$$\mathbf{v} = \mathbf{u}_1 + \dots + \mathbf{u}_s.$$

Here $\mathbf{u}_i = P_i(\mathbf{v}) \in U_i$. From this expansion we derive

$$P_i(f(\mathbf{v})) = P_i(f(\mathbf{u}_1) + \dots + f(\mathbf{u}_s)) = f(\mathbf{u}_i) = f(P_i(\mathbf{v}))$$

We used the inclusion $\mathbf{w}_j = f(\mathbf{u}_j) \in U_j$ that follows from the invariance of the subspace U_j under the action of f . We also used the following properties of projection operators (they follow from (2.11) and $U_i = \text{Im } P_i$, see § 2 above):

$$P_i(\mathbf{w}_j) = \begin{cases} \mathbf{w}_i & \text{for } j = i, \\ \mathbf{0} & \text{for } j \neq i. \end{cases}$$

Since $\mathbf{v}$ is an arbitrary vector of the space V , from the above equality $P_i(f(\mathbf{v})) = f(P_i(\mathbf{v}))$ we derive $f \circ P_i = P_i \circ f$.

Conversely, suppose that the operator f commute with all projection operators $P_1, \dots, P_s$ associated with the expansion $V = U_1 \oplus \dots \oplus U_s$. Let $\mathbf{u}$ be an arbitrary vector of the subspace U_i . Then we denote $\mathbf{w} = f(\mathbf{u})$ and for $\mathbf{w}$ we derive

$$P_i(\mathbf{w}) = P_i(f(\mathbf{u})) = f(P_i(\mathbf{u})) = f(\mathbf{u}) = \mathbf{w}.$$

Remember that P_i projects onto the subspace U_i . Hence, $P_i(\mathbf{w}) \in U_i$. But due to the above equality we find that $P_i(\mathbf{w}) = \mathbf{w} = f(\mathbf{u}) \in U_i$. Thus we have shown that the space U_i is invariant under the action of the operator f . The theorem is completely proved. $\square$

Let's consider a linear operator f in a finite-dimensional linear vector space V and possessing an invariant subspace U . Suppose that $\dim V = n$ and $\dim U = s$. Let's choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ in U and then, if $s < n$, complete this basis up to a basis in V . Denote by $\mathbf{e}_{s+1}, \dots, \mathbf{e}_n$ the complementary vectors. For $j \leq s$ due to the invariance of the subspace U under the action of f we have $f(\mathbf{e}_j) \in U$. Therefore, in the expansions of these vectors

$$f(\mathbf{e}_j) = \sum_{i=1}^s F_j^i \cdot \mathbf{e}_i, \text{ where } j \leq s,$$

the summation index i runs from 1 to s , but not from 1 to n as it should in general case, where we expand an arbitrary vector of V . This means that if we construct the matrix of the operator f in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$, this matrix would

be mounted of blocks with the lower left block in it being zero:

$$F = \left\| \begin{array}{cccccc} \overbrace{F_1^1 & F_2^1 & \dots & F_s^1}^s & F_{s+1}^1 & \dots & F_n^1 \\ F_1^2 & F_2^2 & \dots & F_s^2 & F_{s+1}^2 & \dots & F_n^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ F_1^s & F_2^s & \dots & F_s^s & F_{s+1}^s & \dots & F_n^s \\ 0 & 0 & \dots & 0 & F_{s+1}^{s+1} & \dots & F_n^{s+1} \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & F_{s+1}^n & \dots & F_n^n \end{array} \right\|_s \quad (3.3)$$

Matrices of this form are called *blockwise-triangular* matrices. The upper left diagonal block in the matrix (3.3) coincides with the matrix of restricted operator $f_U: U \rightarrow U$ in the invariant subspace U .

The lower right diagonal block of the matrix (3.3) can also be interpreted in a special way. In order to find this interpretation let's consider the cosets of complementary vectors in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$:

$$\mathbf{E}_1 = \text{Cl}_U(\mathbf{e}_{s+1}), \dots, \mathbf{E}_{n-s} = \text{Cl}_U(\mathbf{e}_n). \quad (3.4)$$

When proving the theorem 7.6 in Chapter I, we have found that these cosets form a basis in the factorspace V/U . Applying the factoroperator $f_{V/U}$ to (3.4), we get

$$\begin{aligned} f_{V/U}(\mathbf{E}_j) &= f_{V/U} \text{Cl}_U(\mathbf{e}_{s+j}) = \text{Cl}_U(f(\mathbf{e}_{s+j})) = \\ &= \sum_{i=1}^s F_{s+j}^i \cdot \text{Cl}_U(\mathbf{e}_i) + \sum_{i=s+1}^n F_{s+j}^i \cdot \text{Cl}_U(\mathbf{e}_i). \end{aligned}$$

The first sum in the above expression is equal to zero since the vectors $\mathbf{e}_1, \dots, \mathbf{e}_s$ belong to U . Then, shifting the index $i + s \rightarrow i$, we find

$$f_{V/U}(\mathbf{E}_j) = \sum_{i=1}^{n-s} F_{s+j}^{s+i} \cdot \mathbf{E}_i.$$

Looking at this formula, we see that the matrix of the factoroperator $f_{V/U}$ in the basis (3.4) coincides with the lower right diagonal block in the matrix (3.3).

THEOREM 3.6. *Let $f: V \rightarrow V$ be a linear operator in a finite-dimensional space and let U be an invariant subspace of this operator. Then the determinant of f is equal to the product of two determinants — the determinant of the restricted operator f_U and that of the factoroperator $f_{V/U}$:*

$$\det f = \det(f_U) \cdot \det(f_{V/U}).$$

The proof of this theorem is immediate from the following fact well-known in the theory of determinants: the determinant of blockwise-triangular matrix is equal to the product of determinants of all its diagonal blocks.

§ 4. Eigenvalues and eigenvectors.

Let $f: V \rightarrow V$ be a linear operator. A nonzero vector $\mathbf{v} \neq \mathbf{0}$ of the space V is called an *eigenvector* of the operator f if $f\mathbf{v} = \lambda \cdot \mathbf{v}$, where $\lambda \in \mathbb{K}$. The number λ is called the *eigenvalue* of the operator f associated with the eigenvector $\mathbf{v}$.

One eigenvalue λ of an operator f can be associated with several or even with infinite number of eigenvectors. But conversely, if an eigenvector is given, the associated eigenvalue λ for this eigenvector is unique. Indeed, from the equality $f\mathbf{v} = \lambda \cdot \mathbf{v} = \lambda' \cdot \mathbf{v}$ and from $\mathbf{v} \neq \mathbf{0}$ it follows that $\lambda = \lambda'$.

Let $\mathbf{v}$ be an eigenvector of the operator $f: V \rightarrow V$. Let's consider the other operator $h_\lambda = f - \lambda \cdot 1$. Then the equation $f\mathbf{v} = \lambda \cdot \mathbf{v}$ can be rewritten as

$$(f - \lambda \cdot 1)\mathbf{v} = \mathbf{0}. \quad (4.1)$$

Hence, $\mathbf{v} \in \text{Ker}(f - \lambda \cdot 1)$. The condition $\mathbf{v} \neq \mathbf{0}$ means that the kernel of this operator is nonzero: $\text{Ker}(f - \lambda \cdot 1) \neq \{\mathbf{0}\}$.

DEFINITION 4.1. A number $\lambda \in \mathbb{K}$ is called an *eigenvalue* of a linear operator $f: V \rightarrow V$ if the subspace $V_\lambda = \text{Ker}(f - \lambda \cdot 1)$ is nonzero. This subspace $V_\lambda = \text{Ker}(f - \lambda \cdot 1) \neq \{\mathbf{0}\}$ is called the *eigenspace* associated with the eigenvalue λ , while any nonzero vector of V_λ is called an *eigenvector* of the operator f associated with the eigenvalue λ .

The collection of all eigenvalues of an operator f is sometimes called the *spectrum* of this operator, while the brunch of mathematics studying the spectra of linear operators is known as the *spectral theory of operators*. The spectral theory of linear operators in finite-dimensional spaces is the most simple one. This is the very theory that is usually studied in the course of linear algebra.

Let $f: V \rightarrow V$ be a linear operator in a finite-dimensional linear vector space V . In order to find the spectrum of this operator we apply the corollary of theorem 1.4. Due to this corollary a number $\lambda \in \mathbb{K}$ is an eigenvalue of the operator f if and only if it satisfies the equation

$$\det(f - \lambda \cdot 1) = 0. \quad (4.2)$$

The equation (4.2) is called the *characteristic equation* of the operator f , its roots are called the *characteristic numbers* of the operator f .

Let $\dim V = n$. Then the determinant in formula (4.2) is equal to the determinant of the square $n \times n$ matrix. The matrix of the operator $h_\lambda = f - \lambda \cdot 1$ is derived from the matrix of the operator f by subtracting λ from each element on the primary diagonal of this matrix:

$$H_\lambda = \begin{vmatrix} F_1^1 - \lambda & F_2^1 & \dots & F_n^1 \\ F_1^2 & F_2^2 - \lambda & \dots & F_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ F_1^n & F_2^n & \dots & F_n^n - \lambda \end{vmatrix}. \quad (4.3)$$

The determinant of the matrix (4.3) is a polynomial of λ :

$$\det(f - \lambda \cdot 1) = (-\lambda)^n + F_1 (-\lambda)^{n-1} + \dots + F_n. \quad (4.4)$$

The polynomial in right hand side of (4.4) is called the *characteristic polynomial* of the operator f . If F is the matrix of the operator f in some basis, then the coefficients $F_1, \dots, F_n$ of characteristic polynomial (4.4) are expressed through the elements of the matrix F . However, note that left hand side of (4.4) is basis independent, therefore, the coefficients $F_1, \dots, F_n$ do not actually depend on the choice of basis. They are scalar invariants of the operator f . The first and the last invariants in (4.4) are the most popular ones:

$$F_1 = \operatorname{tr} f, \quad F_n = \det f.$$

The invariant F_1 is called the *trace* of the operator f . It is calculated through the matrix of this operator according to the following formula:

$$\operatorname{tr} f = \sum_{i=1}^n F_i^i. \quad (4.5)$$

We shall not derive this formula (4.5) since it is well-known in the theory of determinants. We shall only derive the invariance of the trace immediately on the base of formula (1.4) which describes the transformation of the matrix of a linear operator under a change of basis:

$$\sum_{p=1}^n \tilde{F}_p^p = \sum_{i=1}^n \sum_{j=1}^n \left(\sum_{p=1}^n T_i^p S_p^j \right) F_j^i = \sum_{i=1}^n \sum_{j=1}^n \delta_i^j F_j^i = \sum_{i=1}^n F_i^i.$$

Upon substituting (4.4) into (4.2) we see that the characteristic equation (4.2) of the operator f is a polynomial equation of n -th order with respect to λ :

$$(-\lambda)^n + F_1 (-\lambda)^{n-1} + \dots + F_n = 0. \quad (4.6)$$

Therefore we can estimate the number of eigenvalues of the operator f . Any eigenvalue $\lambda \in \mathbb{K}$ is a root of characteristic equation (4.6). However, not any root of the equation (4.6) is an eigenvalue of the operator f . The matter is that a polynomial equation with coefficients in the numeric field $\mathbb{K}$ can have roots in some larger field $\tilde{\mathbb{K}}$ (e. g. $\mathbb{Q} \subset \mathbb{R}$ or $\mathbb{R} \subset \mathbb{C}$). For the characteristic number λ of the operator f to be an eigenvalue of this operator it should belong to $\mathbb{K}$. From the course of general algebra we know that the total number of roots of the equation (4.6) counted according to their multiplicity and including those belonging to the extensions of the field $\mathbb{K}$ is equal to n (see [4]).

THEOREM 4.1. *The number of eigenvalues of a linear operator $f: V \rightarrow V$ equals to the dimension of the space V at most.*

Consider the case $\mathbb{K} = \mathbb{Q}$. The roots of a polynomial equation with rational coefficients are not necessarily rational numbers: the equation $\lambda^2 - 3 = 0$ is an example. In the case of real numbers $\mathbb{K} = \mathbb{R}$ a polynomial equation with real coefficients can also have non-real roots, e. g. the equation $\lambda^2 + \sqrt{3} = 0$. However, the field of complex numbers $\mathbb{K} = \mathbb{C}$ is an exception.

THEOREM 4.2. *An arbitrary polynomial equation of n -th order with complex coefficients has exactly n complex roots counted according to their multiplicity.*

We shall not prove here this theorem referring the reader to the course of general algebra (see [4]). The theorem 4.2 is known as the «basic theorem of algebra», while the property of complex numbers stated in this theorem is called the *algebraic closure* of $\mathbb{C}$, i.e. $\mathbb{C}$ is an algebraically closed numeric field.

DEFINITION 4.2. A numeric field $\mathbb{K}$ is called an *algebraically closed* field if the roots of any polynomial equation with coefficients from $\mathbb{K}$ are again in $\mathbb{K}$.

Certainly, $\mathbb{C}$ is not the unique algebraically closed field. However, in the list of numeric fields $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ that we consider in this book, only the field of complex numbers is algebraically closed.

Let λ be an eigenvalue of a linear operator f . Then λ is a root of the equation (4.6). The multiplicity of this root λ in the equation (4.6) is called the *multiplicity* of the eigenvalue λ .

THEOREM 4.3. *For a linear operator $f: V \rightarrow V$ in a complex linear vector space V the number of its eigenvalues counted according to their multiplicities is exactly equal to the dimension of V .*

This proposition strengthens the theorem 4.1. It is an immediate consequence of the algebraic closure of the field of complex numbers $\mathbb{C}$. In the case $\mathbb{K} = \mathbb{C}$ the characteristic polynomial (4.4) is factorized into a product of terms linear in λ :

$$\det(f - \lambda \cdot 1) = \prod_{i=1}^n (\lambda_i - \lambda). \quad (4.7)$$

For some operators such an expansion can occur in the case $\mathbb{K} = \mathbb{Q}$ or $\mathbb{K} = \mathbb{R}$, however, it is not a typical situation. If $\lambda_1, \dots, \lambda_n$ are understood as characteristic numbers of the operator f , then the formula (4.7) is always valid.

Due to the formula (4.7) we can present the numeric invariants $F_1, \dots, F_n$ of the operator f as elementary symmetric polynomials of its characteristic numbers:

$$F_i = \sigma_i(\lambda_1, \dots, \lambda_n).$$

In particular, for the trace and for the determinant of the operator f we have

$$\operatorname{tr} f = \sum_{i=1}^n \lambda_i, \quad \det f = \prod_{i=1}^n \lambda_i. \quad (4.8)$$

The theory of symmetric polynomials is given in the course of general algebra (see, for example, the book [4]).

THEOREM 4.4. *For any eigenvalue λ of a linear operator $f: V \rightarrow V$ the associated eigenspace V_λ is invariant under the action of f .*

PROOF. The definition 4.1 of an eigenspace V_λ of a linear operator f can be reformulated as $V_\lambda = \{\mathbf{v} \in V : f(\mathbf{v}) = \lambda \cdot \mathbf{v}\}$. Therefore, $\mathbf{v} \in V_\lambda$ implies $f(\mathbf{v}) = \lambda \cdot \mathbf{v} \in V_\lambda$, which proves the invariance of V_λ . $\square$

We know that the set of linear operators in a space V form the algebra $\text{End}(V)$ over the numeric field $\mathbb{K}$. However, this algebra is too big. Let's consider some operator $f \in \text{End}(V)$ and complement it with the identical operator 1. Within the algebra $\text{End}(V)$ we can take positive integer powers of the operator f , we can multiply them by numbers from $\mathbb{K}$, we can add such products, and we can add to them scalar operators obtained by multiplying the identical operator 1 by various numbers from $\mathbb{K}$. As a result we obtain various operators of the form

$$P(f) = \alpha_p \cdot f^p + \dots + \alpha_1 \cdot f + \alpha_0 \cdot 1. \quad (4.9)$$

The set of all operators of the form (4.9) is called the *polynomial envelope* of the operator f ; it is denoted $\mathbb{K}[f]$. This is a subset of $\text{End}(V)$ closed with respect to all algebraic operations in $\text{End}(V)$. Such subsets are used to be called *subalgebras*. It is important to say that the subalgebra $\mathbb{K}[f]$ is commutative, i.e. for any two polynomials P and Q the corresponding operators (4.9) commute:

$$P(f) \circ Q(f) = Q(f) \circ P(f). \quad (4.10)$$

The equality (4.10) is verified by direct calculation. Indeed, let $P(f)$ and $Q(f)$ be two operator polynomials of the form:

$$P(f) = \sum_{i=0}^p \alpha_i \cdot f^i, \quad Q(f) = \sum_{j=0}^q \beta_j \cdot f^j.$$

Here we denote: $f^0 = 1$. This relationship should be treated as the definition of zeroth power of the operator f . Then

$$P(f) \circ Q(f) = \sum_{i=0}^p \sum_{j=0}^q (\alpha_i \beta_j) \cdot f^{i+j} = Q(f) \circ P(f).$$

These calculations prove the relationship (4.10).

THEOREM 4.5. *Let U be an invariant subspace of an operator f . Then it is invariant under the action of any operator from the polynomial envelope $\mathbb{K}[f]$.*

PROOF. Let $\mathbf{u}$ be an arbitrary vector of U . Let's consider the following vectors $\mathbf{u}_0 = \mathbf{u}$, $\mathbf{u}_1 = f(\mathbf{u})$, $\mathbf{u}_2 = f^2(\mathbf{u})$, $\dots$, $\mathbf{u}_p = f^p(\mathbf{u})$. Every next vector in this sequence is obtained by applying the operator f to the previous one: $\mathbf{u}_{i+1} = f(\mathbf{u}_i)$. Therefore, from $\mathbf{u}_0 \in U$ it follows that $\mathbf{u}_1 \in U$ since U is an invariant subspace of f . Then, in turn, we successively obtain $\mathbf{u}_2 \in U$, $\mathbf{u}_3 \in U$, and so on up to $\mathbf{u}_p \in U$. Applying the operator $P(f)$ of the form (4.9) to the vector $\mathbf{u}$, we get

$$P(f) \mathbf{u} = \alpha_p \cdot \mathbf{u}_p + \dots + \alpha_1 \cdot \mathbf{u}_1 + \alpha_0 \cdot \mathbf{u}_0.$$

Hence, due to $\mathbf{u}_i \in U$ we find that $P(f) \mathbf{u} \in U$, which proves the invariance of U under the action of the operator $P(f)$. $\square$

The following fact is curious: if λ is an eigenvalue of the operator f and if $\mathbf{v}$ is an associated eigenvector, then $P(f) \mathbf{v} = P(\lambda) \cdot \mathbf{v}$. Therefore, any eigenvector $\mathbf{v}$ of

the operator f is an eigenvector of the operator $P(f)$. The converse proposition, however, is not true.

Let $\lambda_1, \dots, \lambda_s$ be a set of mutually distinct eigenvalues of the operator f . Let's consider the operators $h_i = f - \lambda_i \cdot \mathbf{1}$, which certainly belong to the polynomial envelope of f . The permutability of any two such operators follows from (4.10). The eigenspace V_{λ_i} of the operator f is determined as the kernel of the operator h_i . According to the definition 4.1, it is nonzero. Moreover, the theorems 4.4 and 4.5 say that V_{λ_i} is invariant under the action of f and of all other operators h_j .

THEOREM 4.6. *Let $\lambda_1, \dots, \lambda_s$ be a set of mutually distinct eigenvalues of the operator $f: V \rightarrow V$. Then the sum of associated eigenspaces $V_{\lambda_1}, \dots, V_{\lambda_s}$ is a direct sum: $V_{\lambda_1} + \dots + V_{\lambda_s} = V_{\lambda_1} \oplus \dots \oplus V_{\lambda_s}$.*

Note that the set of mutually distinct eigenvalues $\lambda_1, \dots, \lambda_s$ of the operator f in this theorem could be the complete set of such eigenvalues, or it could include only a part of such eigenvalues. This makes no difference for the result of the theorem 4.6, it remains valid in either case.

PROOF. Let's denote by W the sum of eigenspaces of the operator f :

$$W = V_{\lambda_1} \oplus \dots \oplus V_{\lambda_s}. \quad (4.11)$$

In order to prove that the sum (4.11) is a direct sum we need to prove that for an arbitrary vector $\mathbf{w} \in W$ the expansion

$$\mathbf{w} = \mathbf{v}_1 + \dots + \mathbf{v}_s, \text{ where } \mathbf{v}_i \in V_{\lambda_i}, \quad (4.12)$$

is unique. For this purpose we consider the operator f_i defined by formula

$$f_i = \prod_{r \neq i}^s h_r.$$

The operator f_i belongs to the polynomial envelope of the operator f and

$$f_i(\mathbf{v}_j) = \left(\prod_{r \neq i}^s (\lambda_j - \lambda_r) \right) \cdot \mathbf{v}_j. \quad (4.13)$$

This follows from $\mathbf{v}_j \in V_{\lambda_j}$, which implies $h_r(\mathbf{v}_j) = (\lambda_j - \lambda_r) \cdot \mathbf{v}_j$. The formula (4.13) means that $f_i(\mathbf{v}_j) = 0$ for all $j \neq i$. Applying the operator f_i to both sides of the expansion (4.12), we get the equality

$$f_i(\mathbf{w}) = \left(\prod_{r \neq i}^s (\lambda_i - \lambda_r) \right) \cdot \mathbf{v}_i.$$

Hence, for the vector $\mathbf{v}_i$ in the expansion (4.12) we derive

$$\mathbf{v}_i = \frac{f_i(\mathbf{w})}{\prod_{r \neq i}^s (\lambda_i - \lambda_r)}. \quad (4.14)$$

The formula (4.14) uniquely determines all summands in the expansion (4.12) if the vector $\mathbf{w} \in W$ is given. This means that the expansion (4.12) is unique and the sum of subspaces (4.11) is a direct sum. $\square$

DEFINITION 4.3. A linear operator $f: V \rightarrow V$ in a linear vector space V is called a *diagonalizable operator* if there is a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in the space V such that the matrix of the operator f is diagonal in this basis.

THEOREM 4.7. An operator $f: V \rightarrow V$ is diagonalizable if and only if the sum of all its eigenspaces coincides with V .

PROOF. Let f be a diagonalizable operator. Then we can choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ such that its matrix F in this basis is diagonal, i.e. only diagonal elements F_i^i of this matrix can be nonzero. Then the relationship (1.2), which determines the matrix F , is written as $f(\mathbf{e}_i) = F_i^i \cdot \mathbf{e}_i$. Hence, each basis vector $\mathbf{e}_i$ is an eigenvector of the operator f , while $\lambda_i = F_i^i$ is its associated eigenvalue. The expansion of an arbitrary vector $\mathbf{v}$ in this base is an expansion by eigenvectors of the operator f . Therefore, having collected together the terms with coinciding eigenvalues in this expansion, we get the expansion

$$\mathbf{v} = \mathbf{v}_1 + \dots + \mathbf{v}_s, \text{ where } \mathbf{v}_i \in V_{\lambda_i}.$$

Since $\mathbf{v}$ is an arbitrary vector of V , this means that $V_{\lambda_1} + \dots + V_{\lambda_s} = V$. The direct proposition of the theorem is proved.

Conversely, suppose that $\lambda_1, \dots, \lambda_s$ is the total set of mutually distinct eigenvalues of the operator f and assume that $V_{\lambda_1} + \dots + V_{\lambda_s} = V$. The theorem 4.6 says that this is a direct sum: $V = V_{\lambda_1} \oplus \dots \oplus V_{\lambda_s} = V$. Therefore, choosing a basis in each eigenspace and joining them together, we get a basis in V (see theorem 6.3 in Chapter I). This is a basis composed by eigenvectors of the operator f , the application of f to each basis vector reduces to multiplying this vector by its associated eigenvalue. Therefore, the matrix F of the operator f in this basis is diagonal. Its diagonal elements coincide with the eigenvalues of the operator f . The theorem is proved. $\square$

Assume that an operator $f: V \rightarrow V$ is diagonalizable and assume that we have chosen a basis where its matrix is diagonal. Then the matrix H_f in formula (4.3) is also diagonal. Hence, we immediately derive the following formula:

$$\det(f - \lambda \cdot \mathbf{1}) = \prod_{i=1}^n (F_i^i - \lambda).$$

Due to this equality we conclude that the characteristic polynomial of a diagonalizable operator is factorized into the product of a linear terms and all roots of characteristic equation belong to the field $\mathbb{K}$ (not to its extension). This means that characteristic numbers of a diagonalizable operator coincide with its eigenvalues. This is a necessary condition for the operator f to be diagonalizable. However, it is not a sufficient condition. Even in the case of algebraically closed field of complex numbers $\mathbb{K} = \mathbb{C}$ there are non-diagonalizable operators in vector spaces over the field $\mathbb{C}$.

§ 5. Nilpotent operators.

DEFINITION 5.1. A linear operator $f: V \rightarrow V$ is called a *nilpotent operator* if for any vector $\mathbf{v} \in V$ there is a positive integer number k such that $f^k(\mathbf{v}) = \mathbf{0}$.

According to the definition 5.1 for any vector v there is an integer number k (depending on $\mathbf{v}$) such that $f^k(\mathbf{v}) = \mathbf{0}$. The choice of such number has no upper bound, indeed, if $m > k$ and $f^k(\mathbf{v}) = \mathbf{0}$ then $f^m(\mathbf{v}) = \mathbf{0}$. This means that there is a minimal positive number $k = k_{\min}$ (depending on $\mathbf{v}$) such that $f^k(\mathbf{v}) = \mathbf{0}$. This minimal number $k_{\min}$ is called the *height* of the vector $\mathbf{v}$ relative to the nilpotent operator f . The height of zero vector is taken to be zero by definition; for any nonzero vector $\mathbf{v}$ its height is greater or equal to the unity. Let's denote the height of $\mathbf{v}$ by $\nu(\mathbf{v})$ and define the number

$$\nu(f) = \max_{\mathbf{v} \in V} \nu(\mathbf{v}). \quad (5.1)$$

For each vector $\mathbf{v} \in V$ its height is finite, but the maximum in (5.1) can be infinite since the number of vectors in a linear vector space usually is infinite.

DEFINITION 5.2. In that case, where the maximum in the formula (5.1) is finite, a nilpotent operator f is called an *operator of finite height* and the number $\nu(f)$ is called the *height* of a nilpotent operator f .

THEOREM 5.1. In a finite-dimensional linear vector space V the height $\nu(f)$ of any nilpotent operator $f: V \rightarrow V$ is finite.

PROOF. Let's choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in V and consider the heights of all basis vectors $\nu(\mathbf{e}_1), \dots, \nu(\mathbf{e}_n)$ with respect to f . Then denote

$$m = \max\{\nu(\mathbf{e}_1), \dots, \nu(\mathbf{e}_n)\}.$$

For an arbitrary vector $\mathbf{v} \in V$ consider its expansion $\mathbf{v} = v^1 \cdot \mathbf{e}_1 + \dots + v^n \cdot \mathbf{e}_n$. Then, applying the operator f^m to $\mathbf{v}$, we find

$$f^m(\mathbf{v}) = \sum_{i=1}^n v^i \cdot f^m(\mathbf{e}_i) = \mathbf{0}. \quad (5.2)$$

Due to the formula (5.2) we see that the heights of all vectors of the space V are restricted by the number m . This means that the height of a nilpotent operator f is finite: $\nu(f) = m < \infty$. $\square$

THEOREM 5.2. If $f: V \rightarrow V$ is a nilpotent operator and if U is an invariant subspace of the operator f , then the restricted operator f_U and the factoroperator $f_{V/U}$ both are nilpotent.

PROOF. Any vector $\mathbf{u}$ of the subspace $U \subset V$ is a vector of V . Therefore, there is an integer number $k > 0$ such that $f^k(\mathbf{u}) = \mathbf{0}$. However, the result of applying the restricted operator f_U to a vector of U coincides with the result of applying the initial operator f to this vector. Hence, we have

$$(f_U)^k \mathbf{u} = f^k(\mathbf{u}) = \mathbf{0}.$$

This proves that f_U is a nilpotent operator. In the case of factoroperator we consider an arbitrary coset Q in the factorspace V/U . Let $Q = \text{Cl}_U(\mathbf{v})$, where v is some fixed vector in Q , and let $k = \nu(v)$ be the height of this vector $\mathbf{v}$ relative to the operator f . Then we can calculate

$$(f_{V/U})^k Q = \text{Cl}_U(f^k(\mathbf{v})) = \mathbf{0}.$$

Now it is clear that the factoroperator $f_{V/U}$ is a nilpotent operator. The theorem is completely proved. $\square$

THEOREM 5.3. *A nilpotent operator f cannot have a nonzero eigenvalue.*

PROOF. Let λ be an eigenvalue of a nilpotent operator f and let $\mathbf{v} \neq \mathbf{0}$ be an associated eigenvector. Then we have $f(\mathbf{v}) = \lambda \cdot \mathbf{v}$. On the other hand, since f is nilpotent, there is a number $k > 0$ such that $f^k(\mathbf{v}) = \mathbf{0}$. Then we derive

$$f^k(\mathbf{v}) = \lambda^k \cdot \mathbf{v} = \mathbf{0}.$$

But $\mathbf{v} \neq \mathbf{0}$, therefore, $\lambda^k = 0$. This is the equation for λ and $\lambda = 0$ its unique root. The theorem is proved. $\square$

It the finite-dimensional case this theorem can be strengthened as follows.

THEOREM 5.4. *In a finite-dimensional space V of the dimension $\dim V = n$ any nilpotent operator f has exactly one eigenvalue $\lambda = 0$ with the multiplicity n .*

PROOF. We shall prove this theorem by induction on $n = \dim V$. In the case $n = 1$ we fix some vector $\mathbf{v} \neq \mathbf{0}$ in V and denote by $k = \nu(\mathbf{v})$ its height. Then $f^k(\mathbf{v}) = \mathbf{0}$ and $f^{k-1}(\mathbf{v}) \neq \mathbf{0}$. This means that $\mathbf{w} = f^{k-1}(\mathbf{v}) \neq \mathbf{0}$ is an eigenvector of f with the eigenvalue $\lambda = 0$ since $f(\mathbf{w}) = f^k(\mathbf{v}) = \mathbf{0} = 0 \cdot \mathbf{w}$. The base of the induction is proved.

Suppose that the theorem is proved for any finite-dimensional space of the dimension less than n and consider a space V of the dimension $n = \dim V$. As above, let's fix some vector $\mathbf{v} \neq \mathbf{0}$ in V and denote by $k = \nu(\mathbf{v})$ its height relative to the operator f . Then $f^k(\mathbf{v}) = \mathbf{0}$ and $\mathbf{w} = f^{k-1}(\mathbf{v}) \neq \mathbf{0}$. Hence, for the nonzero vector $\mathbf{w}$ we get the following series of equalities:

$$f(\mathbf{w}) = f(f^{k-1}(\mathbf{v})) = f^k(\mathbf{v}) = \mathbf{0} = 0 \cdot \mathbf{w}.$$

Hence, $\mathbf{w}$ is an eigenvector of the operator f and $\lambda = 0$ is its associated eigenvalue. Let's consider the eigenspace $U = V_0$ corresponding to the eigenvalue $\lambda = 0$. Let's denote $m = \dim U \neq 0$. The restricted operator f_U is zero, hence, for characteristic polynomial of this operator $f_U = 0$ we derive

$$\det(f_U - \lambda \cdot 1) = (-\lambda)^m.$$

Now, applying the theorem 3.6, we derive the characteristic polynomial of f :

$$\det(f - \lambda \cdot 1) = (-\lambda)^m \det(f_{V/U} - \lambda \cdot 1). \quad (5.3)$$

The factoroperator $f_{V/U}$ is an operator in factorspace V/U whose dimension $n - m$ is less than n . Due to the theorem 5.2 the factoroperator $f_{V/U}$ is nilpotent,

therefore, we can apply the inductive hypothesis to it. Then for its characteristic polynomial of the factoroperator $f_{V/U}$ we get

$$\det(f_{V/U} - \lambda \cdot 1) = (-\lambda)^{n-m}. \quad (5.4)$$

Comparing the above relationships (5.3) and (5.4), we find the characteristic polynomial of the initial operator f :

$$\det(f - \lambda \cdot 1) = (-\lambda)^n.$$

This means that $\lambda = 0$ is the only eigenvalue of the operator f and its multiplicity is $n = \dim V$. The theorem is proved. $\square$

Let $f: V \rightarrow V$ be a linear operator. Consider a vector $\mathbf{v} \in V$ and denote by $k = \nu(\mathbf{v})$ its height respective to the operator f . This vector $\mathbf{v}$ produces the chain of k vectors according to the following formulas:

$$\mathbf{v}_1 = f^{k-1}(\mathbf{v}), \quad \mathbf{v}_2 = f^{k-2}(\mathbf{v}), \quad \dots, \quad \mathbf{v}_k = f^0(\mathbf{v}) = \mathbf{v}. \quad (5.5)$$

The chain vectors (5.5) are related with each other as follows: $\mathbf{v}_i = f(\mathbf{v}_{i-1})$. Let's apply the operator f to each vector in the chain (5.5). Then the first vector $\mathbf{v}_1$ vanished. Applying f to the rest $k-1$ vectors we get another chain:

$$\mathbf{w}_1 = f^{k-1}(\mathbf{v}), \quad \mathbf{w}_2 = f^{k-2}(\mathbf{v}), \quad \dots, \quad \mathbf{w}_{k-1} = f(\mathbf{v}). \quad (5.6)$$

Comparing these two chains (5.5) and (5.6), we see that they are almost the same, but the second chain is shorter. It is obtained from the first one by removing the last vector $\mathbf{v}_k = \mathbf{v}$.

The vector $\mathbf{v}_1$ is called the *side vector* or the *eigenvector* of the chain (5.5). The other vectors are called the *adjoint vectors* of the chain. If the side vectors of two chains are different, then in these two chains there are no coinciding vectors at all. However, there is even stronger result. It is known as the theorem on «linear independence of chains».

THEOREM 5.5. *If the side vectors in several chains of the form (5.5) are linearly independent, then the whole set of vectors in these chains is linearly independent.*

PROOF. We consider s chains of the form (5.5). In order to specify the chain vector we use two indices $\mathbf{v}_{i,j}$. The first index i is the number of chain to which this vector $\mathbf{v}_{i,j}$ belongs, the second index j specifies the number of this vector within the i -th chain. Denote by $k_1, \dots, k_s$ the lengths of our chains. Without loss of generality we can assume that the chains are arranged in the order of decreasing their lengths, i.e. we have the following inequalities:

$$k_1 \geq k_2 \geq \dots \geq k_s \geq 1. \quad (5.7)$$

Let $k = \max\{k_1, \dots, k_s\}$. We shall prove the theorem by induction on k . If $k = 1$ then the lengths of all chains are equal to 1. Therefore, they contain only the side vectors and have no adjoint vectors at all. The proposition of the theorem in this case is obviously true.

Suppose that the theorem is valid for the chains whose lengths are not greater than $k - 1$. For our s chains, whose lengths are restricted by the number k , we consider a linear combination of all their vectors being equal to zero:

$$\sum_{i=1}^s \sum_{j=1}^{k_s} \alpha_{i,j} \cdot \mathbf{v}_{i,j} = \mathbf{0}. \quad (5.8)$$

From this equality we should derive the triviality of the linear combination in its left hand side. Let's apply the operator f to both sides of (5.8) and use the following quite obvious relationships:

$$f(\mathbf{v}_{i,j}) = \begin{cases} \mathbf{0}, & \text{for } j = 1, \\ \mathbf{v}_{i,j-1}, & \text{for } j > 1. \end{cases}$$

If we take into account (5.7), then the result of applying f to (5.8) is written as

$$\sum_{i=1}^s \sum_{j=1}^{k_s} \alpha_{i,j} \cdot f(\mathbf{v}_{i,j}) = \sum_{i=1}^r \sum_{j=2}^{k_r} \alpha_{i,j} \cdot \mathbf{v}_{i,j-1} = \mathbf{0}. \quad (5.9)$$

In typical situation $r = s$. However, sometimes certain chains of vectors can drop from the above sums at all. This happens if a part of chains were of the length 1. In this case $r < s$ and $k_{r+1} = \dots = k_s = 1$. The lengths of all chains in (5.7) cannot be equal to 1 since $k > 1$.

Shifting the index $j + 1 \rightarrow j$ in the last sum we can write (5.9) as follows:

$$\sum_{i=1}^r \sum_{j=1}^{k_r-1} \alpha_{i,j+1} \cdot \mathbf{v}_{i,j} = \mathbf{0}. \quad (5.10)$$

The left side of the relationship (5.10) is again a linear combination of chain vectors. Here we have r chains with the lengths 1 less as compared to original ones in (5.8). Now we can apply the inductive hypothesis, which yields the linear independence of all vectors presented in (5.10). Hence, all coefficients of the linear combination in left hand side of (5.10) are equal to zero. When applied to (5.8) this fact means that the most part of terms in left hand side of this equality do actually vanish. The remainder is written as follows:

$$\sum_{i=1}^s \alpha_{i,1} \cdot \mathbf{v}_{i,1} = \mathbf{0}. \quad (5.11)$$

Now in the linear combination (5.11) we have only the side vectors of initial chains. They are linearly independent by the assumption of the theorem. Therefore, the linear combination (5.11) is also trivial. From triviality of (5.10) and (5.11) it follows that the initial linear combination (5.8) is trivial too. We have completed the inductive step and thus have proved the theorem in whole. $\square$

Let $f: V \rightarrow V$ be a nilpotent operator in a linear vector space V and let $\mathbf{v}$ be a vector of the height $k = \nu(\mathbf{v})$ in V . Consider the chain of vectors (5.5) generated by $\mathbf{v}$ and denote by $U(\mathbf{v})$ the linear span of chain vectors (5.5):

$$U(\mathbf{v}) = \langle \mathbf{v}_1, \dots, \mathbf{v}_k \rangle. \quad (5.12)$$

Due to the theorem 5.5 the subspace $U(\mathbf{v})$ is a finite-dimensional subspace and $\dim U(\mathbf{v}) = k$. The chain vectors (5.5) form a basis in this subspace (5.12). The following relationships are derived directly from the definition of the chain (5.5):

$$\begin{aligned} f(\mathbf{v}_1) &= \mathbf{0}, \\ f(\mathbf{v}_2) &= \mathbf{v}_1, \\ &\dots\dots\dots \\ f(\mathbf{v}_k) &= \mathbf{v}_{k-1}. \end{aligned} \tag{5.13}$$

Due to (5.13) the subspace (5.12) is invariant under the action of the operator f . Hence, we can consider the restricted operator $f_{U(\mathbf{v})}$ and, using (5.13), we can find the matrix of this restricted operator in the chain basis $\mathbf{v}_1 \dots, \mathbf{v}_k$:

$$J_k(0) = \left\| \begin{array}{cccc} 0 & 1 & & 0 \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{array} \right\|. \tag{5.14}$$

A matrix of the form (5.14) is called a *Jordan block* or a *Jordan cage* of a nilpotent operator. Its primary diagonal is filled with zeros. The upward next diagonal parallel to the primary one is filled with unities. All other space in the matrix (5.14) is filled with zeros again. The matrix (5.14) is a square $k \times k$ matrix, if $k = 1$, this matrix degenerates and becomes purely zero matrix with the only element: $J_1(0) = \|0\|$.

Let $f: V \rightarrow V$ again be a nilpotent operator. We continue to study vector chains of the form (5.5). For this purpose let's consider the following subspaces:

$$U_k = \text{Ker } f \cap \text{Im } f^{k-1}. \tag{5.15}$$

If $\mathbf{u} \in U_k$, then $\mathbf{u} \in \text{Im } f^{k-1}$. Therefore, $\mathbf{u} = f^{k-1}(\mathbf{v})$ for some vector $\mathbf{v}$. This means that $\mathbf{u}$ is a chain vector in a chain of the form (5.5). From the condition $\mathbf{u} \in \text{Ker } f$ we derive $f(\mathbf{u}) = f^k(\mathbf{v}) = \mathbf{0}$. Hence, $\mathbf{v}$ is a vector of the height k and $\mathbf{u}$ is a side vector in the chain (5.5) initiated by the vector $\mathbf{v}$. For the subspaces (5.15) we have the sequence of inclusions

$$V_0 = U_1 \supseteq U_2 \supseteq \dots \supseteq U_k \supseteq \dots, \tag{5.16}$$

where $V_0 = \text{Ker } f$ is the eigenspace corresponding to the unique eigenvalue $\lambda = 0$ of nilpotent operator f . The inclusions (5.16) follow from the fact that any chain (5.5) of the length k with the side vector $\mathbf{u} = f^{k-1}(\mathbf{v})$ can be treated as a chain of the length $k-1$ by dropping the k -th vector $\mathbf{v}_k = \mathbf{v}$ (see (5.5) and (5.6)). Then for the vector $\mathbf{v}' = f(\mathbf{v})$ we have $\mathbf{u} = f^{k-2}(\mathbf{v}')$. This yields the inclusion of subspaces $U_k \subset U_{k-1}$ for $k > 1$.

In a finite-dimensional space V the height of any vector $\mathbf{v} \in V$ is restricted by the height of the nilpotent operator f itself:

$$\nu(\mathbf{v}) \leq \nu(f) = m < \infty$$

(see theorem 5.1). Therefore $U_{m+1} = \{\mathbf{0}\}$. Hence, the sequence of inclusions (5.16)

terminates on m -th step, i. e. we have a finite sequence of inclusions:

$$V_0 = U_1 \supseteq U_2 \supseteq \dots \supseteq U_m \supseteq \{\mathbf{0}\}. \quad (5.17)$$

Sequences of mutually enclosed subspaces of the form (5.16) or (5.17) are called *flags*, while each particular subspace in a flag is called a *flag subspace*.

THEOREM 5.6. *For any nilpotent operator f in a finite-dimensional space V there is a basis in V composed by chain vectors of the form (5.5). Such a basis is called a *canonic basis* or a *Jordan basis* of a nilpotent operator f .*

PROOF. The proof of the theorem is based on the fact that the flag (5.17) is finite. We choose a basis in the smallest subspace U_m . Then we complete it up to a basis in U_{m-1} , in U_{m-2} , and so on backward along the sequence (5.17). As a result we construct a basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ in $V_0 = \text{Ker } f$. Note that each vector in such a basis is a side vector of some chain of the form (5.5). For basis vectors of the subspace U_m the lengths of such chains are equal to m . For the complementary vectors from U_{m-1} their chains are of the length $m-1$ and further the length of chains decreases step by step until the unity for the complementary vectors in largest subspace $U_1 = V_0$.

Let's join together all vectors of the above chains and let's enumerate them by means of double indices: $\mathbf{e}_{i,j}$. Here i is the number of the chain and j is the individual number of the vector within i -th chain. Then

$$\mathbf{e}_1 = \mathbf{e}_{1,1}, \dots, \mathbf{e}_s = \mathbf{e}_{s,1}.$$

Now let's prove that the set of all vectors from the above chains form a basis in V . The linear independence of this set of vectors follows from the theorem 5.5. We only have to prove that an arbitrary vector $\mathbf{v} \in V$ can be represented as a linear combination of chain vectors $\mathbf{e}_{i,j}$. We shall prove this fact by induction on the height of the vector $\mathbf{v}$.

If $k = \nu(\mathbf{v}) = 1$, then $\mathbf{v} \in \text{Ker } f = V_0$. In this case $\mathbf{v}$ is expanded in the basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ of the subspace V_0 . This is the base of induction.

Now suppose that any vector of the height less than k can be represented as a linear combination of chain vectors $\mathbf{e}_{i,j}$. Let's take a vector $\mathbf{v}$ of the height k and denote $\mathbf{u} = f^{k-1}(\mathbf{v})$. Then $f(\mathbf{u}) = \mathbf{0}$. This means that $\mathbf{u}$ is a side vector in a chain of the length k initiated by the vector $\mathbf{v}$. Therefore, $\mathbf{u}$ is an element of the subspace U_k (see formula (5.15)); this vector can be expanded in the basis of the subspace U_k , which we have constructed above:

$$\mathbf{u} = \sum_{i=1}^r \alpha_i \cdot \mathbf{e}_i. \quad (5.18)$$

Note that in the expansion (5.18) we have only a part of vectors $\mathbf{e}_1, \dots, \mathbf{e}_s$, namely, we have only those of them that belongs to U_k and, hence, are side vectors in the chains of the length not less than k . Therefore, we can write $\mathbf{e}_i = f^{k-1}(\mathbf{e}_{i,k})$ for $i = 1, \dots, r$. Substituting these expressions into (5.18), we obtain

$$f^{k-1}(\mathbf{v}) = \sum_{i=1}^r \alpha_i \cdot f^{k-1}(\mathbf{e}_{i,k}). \quad (5.19)$$

By means of the coefficients of the expansion (5.19) we determine the vector $\mathbf{v}'$:

$$\mathbf{v}' = \mathbf{v} - \sum_{i=1}^r \alpha_i \cdot \mathbf{e}_{i,k}. \quad (5.20)$$

Applying the operator f^{k-1} to $\mathbf{v}'$ and taking into account (5.19), we find

$$f^{k-1}(\mathbf{v}') = f^{k-1}(\mathbf{v}) - \sum_{i=1}^r \alpha_i \cdot \mathbf{f}^{k-1}(\mathbf{e}_{i,k}) = \mathbf{0}.$$

Hence, the height of the vector $\mathbf{v}'$ is less than k and we can apply the inductive hypothesis to it. This means that $\mathbf{v}'$ can be represented as a linear combination of chain vectors $\mathbf{e}_{i,j}$. But $\mathbf{v}$ is expressed through $\mathbf{v}'$ as follows:

$$\mathbf{v} = \mathbf{v}' + \sum_{i=1}^r \alpha_i \cdot \mathbf{e}_{i,k}.$$

Then $\mathbf{v}$ can also be expressed as a linear combination of chain vectors $\mathbf{e}_{i,j}$. The inductive step is completed and the theorem in whole is proved. $\square$

In the basis composed by chain vectors, the existence of which was proved in theorem 5.6, the matrix of nilpotent operator f has the following form:

$$F = \left\| \begin{array}{cccc} J_{k_1}(0) & & & \\ & J_{k_2}(0) & & \\ & & \ddots & \\ & & & J_{k_s}(0) \end{array} \right\|. \quad (5.21)$$

The matrix (5.21) is blockwise-diagonal, its diagonal blocks are Jordan cages of the form (5.14), all other space in this matrix is filled with zeros. It is easy to understand this fact. Indeed, each chain with the side vector $\mathbf{e}_i$ produces the invariant subspace $U(v)$ of the form (5.12), where $\mathbf{v} = \mathbf{e}_{i,k_i}$. Due to the theorem 5.6 the space V is the direct sum of such invariant subspaces:

$$V = U(\mathbf{e}_{1,k_1}) \oplus \dots \oplus U(\mathbf{e}_{s,k_s}).$$

The matrix (5.21) is called a *Jordan form* of the matrix of a nilpotent operator. The theorem 5.6 is known as *the theorem on bringing the matrix of a nilpotent operator to a canonic Jordan form*. If the chain basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ is constructed strictly according to the proof of the theorem 5.6, then Jordan cages are arranged in the order of decreasing sizes:

$$k_1 \geq k_2 \geq \dots \geq k_s.$$

However, the permutation of vectors $\mathbf{e}_1, \dots, \mathbf{e}_s$ can change this order, and this usually happens in practice.

THEOREM 5.7. *The height of a nilpotent operator f in a finite-dimensional space V is less or equal to the dimension $n = \dim V$ of this space and $f^n = 0$.*

PROOF. Above in proving the theorem 5.1 we noted that the height $\nu(f)$ of a nilpotent operator f coincides with the greatest height of basis vectors. Due to the theorem 5.6 now we can choose the chain basis. The height of a chain vector is not greater than the length of the chain (5.5) to which it belongs. Therefore, the height of basis vectors in a chain basis is not greater than the number of vectors in such a basis. This yields $\nu(f) \leq n = \dim V$. The height of an arbitrary vector $\mathbf{v}$ of V is not greater than the height of the operator f . Therefore, $f^n(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in V$. This means that $f^n = 0$. The theorem is proved. $\square$

§ 6. Root subspaces. Two theorems on the sum of root subspaces.

DEFINITION 6.1. The *root subspace* of a linear operator $f: V \rightarrow V$ corresponding to its eigenvalue λ is the set

$$V(\lambda) = \{\mathbf{v} \in V : \exists k ((k \in \mathbb{N}) \ \& \ ((f - \lambda \cdot 1)^k \mathbf{v} = \mathbf{0}))\}$$

that consist of vectors vanishing under the action of some positive integer power of the operator $h_\lambda = f - \lambda \cdot 1$.

For each positive integer k we define the subspace $V(k, \lambda) = \text{Ker}(h_\lambda)^k$. For $k = 1$ the subspace $V(1, \lambda)$ coincides with the eigenspace V_λ . Note that $(h_\lambda)^k \mathbf{v} = \mathbf{0}$ implies $(h_\lambda)^{k+1} \mathbf{v} = \mathbf{0}$. Therefore we have the sequence of inclusions

$$V(1, \lambda) \subseteq V(2, \lambda) \subseteq \dots \subseteq V(k, \lambda) \subseteq \dots \quad (6.1)$$

It is easy to see that all subspaces in the sequence (6.1) are enclosed into the root subspace $V(\lambda)$. Moreover, $V(\lambda)$ is the union of the subspaces (6.1):

$$V(\lambda) = \bigcup_{k=1}^{\infty} V(k, \lambda) = \sum_{k=1}^{\infty} V(k, \lambda). \quad (6.2)$$

In this case the sum of subspaces the sum of subspaces $V(k, \lambda)$ coincides with their union. Indeed, let $\mathbf{v}$ be a vector of the sum of subspaces $V(k, \lambda)$. Then

$$\mathbf{v} = \mathbf{v}_{k_1} + \dots + \mathbf{v}_{k_s}, \text{ where } \mathbf{v}_{k_s} \in V(k_s, \lambda). \quad (6.3)$$

Let $k = \max\{k_1, \dots, k_s\}$, then from the sequence of inclusions (6.1) we derive $\mathbf{v}_{k_i} \in V(k, \lambda)$. Therefore the vector (6.3) belongs to $V(k, \lambda)$, hence, it belongs to the union of all subspaces $V(k, \lambda)$.

The proof of coincidence of the sum and the union in (6.2) is based on the inclusions (6.1). Therefore, we have proved the more general theorem.

THEOREM 6.1. *The sum of a growing sequence of mutually enclosed subspaces coincides with their union.*

The theorem 6.1 shows that the set $V(\lambda)$ in definition 6.1 is actually a subspace in V . This subspace is nonzero since it comprises the eigenspace V_λ as a subset.

THEOREM 6.2. *A root subspace $V(\lambda)$ of an operator f is invariant under the action of f and of all operators from its polynomial envelope $P(f)$.*

PROOF. Let $\mathbf{v} \in V(\lambda)$. Then there exists a positive integer number k such that $(h_\lambda)^k \mathbf{v} = \mathbf{0}$. Let's consider the vector $\mathbf{w} = f(\mathbf{v})$. For this vector we have

$$(h_\lambda)^k \mathbf{w} = (h_\lambda)^k \circ f \mathbf{v} = f \circ (h_\lambda)^k \mathbf{v} = f((h_\lambda)^k(\mathbf{v})) = \mathbf{0}.$$

Here we used the permutability of the operators h_λ and f , it follows from the inclusion $h_\lambda \in P(f)$. Due to the above equality we have $\mathbf{w} = f(\mathbf{v}) \in V(\lambda)$. The invariance of $V(\lambda)$ under the action of f is proved. Its invariance under the action of operators from $P(f)$ now follows from the theorem 4.5. $\square$

THEOREM 6.3. *Let λ and μ be two eigenvalues of a linear operator $f: V \rightarrow V$. Then the restriction of the operator $h_\lambda = f - \lambda \cdot 1$ to the root subspace $V(\mu)$ is*

- (1) *a bijective operator if $\mu \neq \lambda$;*
- (2) *a nilpotent operator if $\mu = \lambda$.*

PROOF. Let's prove the first proposition of the theorem. We already know that the subspace $V(\mu)$ is invariant under the action of h_λ . For the sake of convenience we denote by $h_{\lambda,\mu}$ the restriction of h_λ to the subspace $V(\mu)$. This is an operator acting from $V(\mu)$ to $V(\mu)$. Let's find its kernel:

$$\text{Ker } h_{\lambda,\mu} = \{\mathbf{v} \in V(\mu): h_\lambda(\mathbf{v}) = \mathbf{0}\} = \text{Ker } h_\lambda \cap V(\mu).$$

The kernel of the operator h_λ by definition coincides with the eigenspace V_λ . Therefore, $\text{Ker } h_{\lambda,\mu} = V_\lambda \cap V(\mu)$.

Let $\mathbf{v}$ be an arbitrary vector of the kernel $\text{Ker } h_{\lambda,\mu}$. Due to the above result $\mathbf{v}$ belongs to V_λ . Therefore, we have the equality

$$f(\mathbf{v}) = \lambda \cdot \mathbf{v}. \quad (6.4)$$

Simultaneously, we have the other condition $\mathbf{v} \in V(\mu)$ which means that there exists some integer number $k > 0$ such that

$$(h_\mu)^k \mathbf{v} = (f - \mu \cdot 1)^k \mathbf{v} = \mathbf{0}. \quad (6.5)$$

From (6.4) we get $h_\mu(\mathbf{v}) = f(\mathbf{v}) - \mu \cdot \mathbf{v} = (\lambda - \mu) \cdot \mathbf{v}$. Combining this equality with (6.5), we obtain the following equality for $\mathbf{v}$:

$$(h_\mu)^k \mathbf{v} = (\lambda - \mu)^k \cdot \mathbf{v} = \mathbf{0}.$$

Therefore, if $\lambda \neq \mu$, we immediately get $\mathbf{v} = \mathbf{0}$, which means that $\text{Ker } h_{\lambda,\mu} = \{\mathbf{0}\}$. Hence, in the case $\lambda \neq \mu$ the operator $h_{\lambda,\mu}: V(\mu) \rightarrow V(\mu)$ is injective. The surjectivity of this operator and, hence, its bijectivity follows from its injectivity due to the theorem 1.3.

Now let's prove the second proposition of the theorem. In this case $\mu = \lambda$, therefore, we consider the operator $h_{\lambda,\lambda}$ being the restriction of h_λ to the subspace $V(\lambda)$. Note that $h_{\lambda,\lambda} \mathbf{v} = h_\lambda \mathbf{v}$ for all $\mathbf{v} \in V(\lambda)$. Therefore, from the definition of a root subspace we conclude that for any vector $\mathbf{v} \in V(\lambda)$ there is a positive integer

number k such that $(h_{\lambda,\lambda})^k \mathbf{v} = (f - \lambda \cdot 1)^k \mathbf{v} = \mathbf{0}$. This equality means that $h_{\lambda,\lambda}$ is a nilpotent operator in $V(\lambda)$. The theorem is proved. $\square$

THEOREM 6.4. *Let $\lambda_1, \dots, \lambda_s$ be a set of mutually distinct eigenvalues of a linear operator $f: V \rightarrow V$. Then the sum of corresponding root subspaces is a direct sum: $V(\lambda_1) + \dots + V(\lambda_s) = V(\lambda_1) \oplus \dots \oplus V(\lambda_s)$.*

PROOF. The proof of this theorem is similar to that of theorem 4.6. Denote by W the sum of subspaces specified in the theorem:

$$W = V(\lambda_1) + \dots + V(\lambda_s). \quad (6.6)$$

In order to prove that the sum (6.6) is a direct sum we should prove the uniqueness of the following expansion for an arbitrary vector $\mathbf{w} \in W$:

$$\mathbf{w} = \mathbf{v}_1 + \dots + \mathbf{v}_s, \text{ where } \mathbf{v}_i \in V(\lambda_i). \quad (6.7)$$

Consider another expansion of the same sort for the same vector $\mathbf{w}$:

$$\mathbf{w} = \tilde{\mathbf{v}}_1 + \dots + \tilde{\mathbf{v}}_s, \text{ where } \tilde{\mathbf{v}}_i \in V(\lambda_i). \quad (6.8)$$

Then let's subtract the second expansion from the first one and for the sake of brevity denote $\mathbf{w}_i = (\mathbf{v}_i - \tilde{\mathbf{v}}_i) \in V(\lambda_i)$. As a result we get

$$\mathbf{w}_1 + \dots + \mathbf{w}_s = \mathbf{0}. \quad (6.9)$$

Denote $h_r = f - \lambda_r \cdot 1$. According to the definition of the root subspace $V(\lambda_r)$, for any vector $\mathbf{w}_r$ in the expansion (6.9) there is some positive integer number k_r such that $(h_r)^{k_r} \mathbf{w}_r = \mathbf{0}$. We use this fact and define the operators

$$f_i = \prod_{r \neq i}^s (h_r)^{k_r}. \quad (6.10)$$

Due to the permutability of the operators $\mathbf{h}_1, \dots, \mathbf{h}_s$ belonging to the polynomial envelope of the operator f and due to the equality $(h_r)^{k_r} \mathbf{w}_r = \mathbf{0}$ we get

$$f_i(\mathbf{w}_j) = \mathbf{0} \text{ for all } j \neq i.$$

Let's apply the operator (6.10) to both sides of the equality (6.9). Then all terms in the sum in left hand side of this equality do vanish, except for i -th term only. This yields $f_i(\mathbf{w}_i) = \mathbf{0}$. Let's write this equality in expanded form:

$$\left(\prod_{r \neq i}^s (h_r)^{k_r} \right) \mathbf{w}_i = \mathbf{0} \quad (6.11)$$

The vector $\mathbf{w}_i$ belongs to the root space $V(\lambda_i)$, which is invariant under the action of all operators h_r in (6.11). Therefore we can replace the operators h_r in (6.11) by their restrictions $h_{r,i}$ to the subspace $V(\lambda_i)$:

$$\left(\prod_{r \neq i}^s (h_{r,i})^{k_r} \right) \mathbf{w}_i = \mathbf{0}. \quad (6.12)$$

According to the theorem 6.3, the restricted operators $h_{r,i}$ are bijective if $r \neq i$. The product (the composition) of bijective operators is bijective. We also know that applying a bijective operator to nonzero vector we would get a nonzero result. Therefore, (6.12) implies $\mathbf{w}_i = \mathbf{0}$. Then $\mathbf{v}_i = \tilde{\mathbf{v}}_i$ and the expansions (6.7) and (6.8) do coincide. The uniqueness of the above expansion (6.7) and the theorem in whole are proved. $\square$

THEOREM 6.5. *Let f be a linear operator in a finite-dimensional space V over the field $\mathbb{K}$ and suppose that its characteristic polynomial factorizes into a product of linear terms in $\mathbb{K}$. Then the sum of all root subspaces of the operator f is equal to V , i.e. $V(\lambda_1) \oplus \dots \oplus V(\lambda_s) = V$, where $\lambda_1, \dots, \lambda_s$ is the set of all mutually distinct eigenvalues of the operator f .*

PROOF. Since $\lambda_1, \dots, \lambda_s$ is the set of all mutually distinct eigenvalues of the operator f , for its characteristic polynomial we get

$$\det(f - \lambda \cdot 1) = \prod_{i=1}^s (\lambda_i - \lambda)^{n_i}.$$

According to the hypothesis of theorem, it is factorized into a product of linear polynomials of the form $\lambda_i - \lambda$, where λ_i is an eigenvalue of f and n_i is the multiplicity of this eigenvalue. Let's denote by W the total sum of all root subspaces of the operator f , we know that this is a direct sum (see theorem 6.4):

$$W = V(\lambda_1) \oplus \dots \oplus V(\lambda_s).$$

The root subspaces are nonzero, hence, $W \neq \{\mathbf{0}\}$.

Further proof is by contradiction. Assume that the proposition of the theorem is false and $W \neq V$. The subspace W is invariant under the action of f as a sum of invariant subspaces $V(\lambda_i)$ (see theorem 3.2). Due to the theorem 4.5 it is invariant under the action of the operator $h_\lambda = f - \lambda \cdot 1$ as well. Let's apply the theorem 3.5 to the operator h_λ . This yields

$$\det(f - \lambda \cdot 1) = \det(f_W - \lambda \cdot 1) \det(f_{V/W} - \lambda \cdot 1). \quad (6.13)$$

Here we took into account that $1_W = 1$ and $1_{V/W} = 1$, we also used the theorem 3.4. The characteristic polynomial of the operator f is the product of characteristic polynomial of restricted operator f_W and that of factoroperator $f_{V/W}$. The left hand side of (6.13) factorizes into a product of linear polynomials in $\mathbb{K}$, therefore, each of the polynomials in right hand side of (6.13) should do the same. Let λ_q be one of the eigenvalues of the factoroperator $f_{V/W}$ and let $Q \in V/W$ be the corresponding eigenvector. Due to (6.13) the number λ_q is in the list $\lambda_1, \dots, \lambda_s$ of eigenvalues of the operator f . Due to our assumption $W \neq V$ we conclude that the factorspace V/W is nontrivial: $V/W \neq \{\mathbf{0}\}$, and the coset Q is not zero. Suppose that $\mathbf{v} \in Q$ is a representative of this coset Q . Since $Q \neq \mathbf{0}$, we have $\mathbf{v} \notin W$. The coset Q is an eigenvector of the factoroperator $f_{V/W}$, therefore, it should satisfy the following equality:

$$(f_{V/W} - \lambda_q \cdot 1)Q = \text{Cl}_W((f - \lambda_q \cdot 1)\mathbf{v}) = \mathbf{0}. \quad (6.14)$$

Let's denote $h_r = f - \lambda_r \cdot 1$ for all $r = 1, \dots, s$ (we have already used this notation in proving the previous theorem). The relationship (6.14) means that

$$(f - \lambda_q \cdot 1) \mathbf{v} = h_q(\mathbf{v}) = \mathbf{w} \in W. \quad (6.15)$$

From the expansion $W = V(\lambda_1) \oplus \dots \oplus V(\lambda_s)$ for the vector $\mathbf{w}$, which arises in formula (6.15), we get the expansion

$$h_q(\mathbf{v}) = \mathbf{w} = \mathbf{v}_1 + \dots + \mathbf{v}_s, \text{ where } \mathbf{v}_i \in V(\lambda_i). \quad (6.16)$$

Let's consider the restriction of the operator h_q to the root subspace $V(\lambda_i)$, this restriction is denoted $h_{q,i}$ (see the proof of theorem 6.4). Due to the theorem 6.3 we know that the operators $h_{q,i} : V(\lambda_i) \rightarrow V(\lambda_i)$ are bijective for all $i \neq q$. Therefore, for all $\mathbf{v}_1, \dots, \mathbf{v}_s$ in (6.16) other than $\mathbf{v}_q$ we can find $\tilde{\mathbf{v}}_i \in V(\lambda_i)$ such that $\mathbf{v}_i = h_{q,i}(\tilde{\mathbf{v}}_i)$. Let's substitute these expressions into (6.16). Then we get

$$\mathbf{w} = h_q(\mathbf{v}) = \mathbf{v}_q + \sum_{i \neq q}^s h_{q,i}(\tilde{\mathbf{v}}_i). \quad (6.17)$$

Relying upon this formula (6.17), we define the new vector $\tilde{\mathbf{v}}_q$:

$$\tilde{\mathbf{v}}_q = \mathbf{v} - \sum_{i \neq q}^s \tilde{\mathbf{v}}_i. \quad (6.18)$$

For this vector from (6.17) we derive $h_q(\tilde{\mathbf{v}}_q) = \mathbf{v}_q \in V(\lambda_q)$. Due to the definition of the root subspace $V(\lambda_q)$ there exists a positive integer number k such that $(h_q)^k \mathbf{v}_q = \mathbf{0}$. Hence, $(h_q)^{k+1} \tilde{\mathbf{v}}_q = \mathbf{0}$ and, therefore, $\tilde{\mathbf{v}}_q \in V(\lambda_q)$. Returning back to the formula (6.18), we derive

$$\mathbf{v} = \sum_{i=1}^s \tilde{\mathbf{v}}_i, \text{ where } \mathbf{v}_i \in V(\lambda_i). \quad (6.19)$$

From the formula (6.19) and from the expansion $W = V(\lambda_1) \oplus \dots \oplus V(\lambda_s)$ it follows that $\mathbf{v} \in W$, but this contradicts to our initial choice $\mathbf{v} \notin W$, which was possible due to the assumption $W \neq V$. Hence, $W = V$. The theorem is proved. $\square$

§ 7. Jordan basis of a linear operator. Hamilton-Cayley theorem.

Let $f: V \rightarrow V$ be a linear operator in finite-dimensional linear vector space V . Suppose that V is expanded into the sum of root subspaces of the operator f :

$$V = V(\lambda_1) \oplus \dots \oplus V(\lambda_s). \quad (7.1)$$

Let's denote $h_i = f - \lambda_i \cdot 1$. Then denote by $h_{i,j}$ the restriction of h_i to $V(\lambda_j)$. According to the theorem 6.3, the restriction $h_{i,i}$ is a nilpotent operator in i -th root subspace $V(\lambda_i)$. Therefore, in $V(\lambda_i)$ we can choose a canonic Jordan basis for this operator (see theorem 5.6). The matrix of the operator $h_{i,i}$ in canonic

Jordan basis is a matrix of the form (5.21) composed by diagonal blocks, where each diagonal block is a matrix of the form (5.14).

DEFINITION 7.1. A *Jordan normal basis* of an operator $f: V \rightarrow V$ is a basis composed by canonic Jordan bases of nilpotent operators $h_{i,i}$ in the root subspaces $V(\lambda_i)$ of the operator f .

Note that an operator f in a finite dimensional space V possesses a Jordan normal basis if and only if there V is expanded into the sum of root subspaces of the operator f , i.e. if we have (7.1). The theorem 6.5 yields a sufficient condition for the existence of a Jordan normal basis of a linear operator.

Suppose that an operator f in a finite-dimensional linear vector space V possesses a Jordan normal basis. The subspaces $V(\lambda_i)$ in (7.1) are invariant with respect to f . Let's denote by f_i the restriction of f to $V(\lambda_i)$. The matrix of the operator f in a Jordan normal basis is a blockwise-diagonal matrix:

$$F = \begin{pmatrix} F_1 & & & \\ & F_2 & & \\ & & \ddots & \\ & & & F_s \end{pmatrix}, \quad (7.2)$$

The diagonal blocks F_i in (7.2) are determined by operators f_i . Note that the operators f_i and $h_{i,i}$ are related to each other by the equality $f_i = h_{i,i} + \lambda_i \cdot 1$. Therefore, F_i is also a blockwise-diagonal matrix:

$$F_i = \begin{pmatrix} J_{k_1}(\lambda_i) & & & \\ & J_{k_2}(\lambda_i) & & \\ & & \ddots & \\ & & & J_{k_r}(\lambda_i) \end{pmatrix}. \quad (7.3)$$

The number of diagonal blocks in (7.3) is determined the number of chains in a canonic Jordan basis of the nilpotent operator $h_{i,i}$, while these diagonal blocks themselves are matrices of the following form:

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}. \quad (7.4)$$

A matrix of the form (7.4) is called a *Jordan block* or a *Jordan cage* with λ on the diagonal. This is square $k \times k$ matrix; if $k = 1$ this matrix degenerates and becomes a matrix with the single element $J_1(\lambda) = \|\lambda\|$.

The matrix of an operator f in a Jordan normal base presented by the relationships (7.2), (7.3), and (7.4) is called a *Jordan normal form* of the matrix of this operator. The problem of constructing a Jordan normal basis for a linear operator f and thus finding the Jordan normal form F of its matrix is known as the problem of *bringing the matrix of a linear operator to a Jordan normal form*.

If the matrix of a linear operator can be brought to a Jordan normal form, this fact has several important consequences. Note that a matrix of the form (7.4) is upper-triangular. Hence, (7.3) and (7.2) all are upper-triangular matrices. The entries on the diagonal of (7.2) are the eigenvalues of the operator f , the i -th eigenvalue λ_i being presented n_i times, where $n_i = \dim V(\lambda_i)$. From the course of algebra we know that the determinant of an upper-triangular matrix is equal to the product of all its diagonal elements. Therefore, the characteristic polynomial of an operator possessing a Jordan normal basis is given by the formula

$$\det(f - \lambda \cdot 1) = \prod_{i=1}^s (\lambda_i - \lambda)^{n_i}. \quad (7.5)$$

THEOREM 7.1. *The matrix of a linear operator f in a finite-dimensional linear vector space V over a numeric field $\mathbb{K}$ can be brought to a Jordan normal form if and only if its characteristic polynomial factorizes into the product of linear polynomials in the field $\mathbb{K}$.*

PROOF. The necessity of the condition formulated in the theorem 7.1 is immediate from (7.5); the sufficiency is provided by the theorems 5.6 and 6.5. $\square$

In the case of the field of complex numbers $\mathbb{C}$ any polynomial factorizes into a product of linear terms. Therefore, the matrix of any linear operator in a complex linear vector space can be brought to a Jordan normal form.

THEOREM 7.2. *The multiplicity of an eigenvalue λ of a linear operator f in a finite-dimensional linear vector space V is equal to the dimension of the corresponding root subspace $V(\lambda)$.*

For the operator f , the characteristic polynomial of which factorizes into the product of linear terms, the proposition of theorem 7.2 immediately follows from the formula (7.5). However, this fact is valid also in the case of partial factorization. Such a case can be reduced to the case of complete factorization by means of the field extension technique. We do not consider the field extension technique in this book. But it is worth to note that the complete proof of the following Hamilton-Cayley theorem is also based on that technique.

THEOREM 7.3. *Let $P(\lambda)$ be the characteristic polynomial of a linear operator f in a finite-dimensional space V . Then $P(f) = 0$.*

PROOF. We shall prove the Hamilton-Cayley theorem for the case where the characteristic polynomial $P(\lambda)$ factorizes into the product of linear terms:

$$P(\lambda) = \prod_{i=1}^s (\lambda_i - \lambda)^{n_i}. \quad (7.6)$$

Denote $h_i := f - \lambda_i \cdot 1$ and denote by $h_{i,j}$ the restriction of h_i to the root subspace $V(\lambda_j)$. Then from the formula (7.6) we derive

$$P(f) = \prod_{i=1}^s (h_i)^{n_i}.$$

Let's apply $P(f)$ to an arbitrary vector $\mathbf{v} \in V$. Due to the theorem (6.5) we can expand $\mathbf{v}$ into a sum $\mathbf{v} = \mathbf{v}_1 + \dots + \mathbf{v}_s$, where $\mathbf{v}_i \in V(\lambda_i)$. Therefore, we have

$$P(f)\mathbf{v} = P(f)\mathbf{v}_1 + \dots + P(f)\mathbf{v}_s. \quad (7.7)$$

The root subspace $V(\lambda_j)$ is invariant under the action of the operators h_i . Then

$$P(f)\mathbf{v}_j = \prod_{i=1}^s (h_i)^{n_i} \mathbf{v}_j = \prod_{i=1}^s (h_{i,j})^{n_i} \mathbf{v}_j.$$

Using permutability of the operators h_i and their restrictions $h_{i,j}$, we can bring the above expression for $P(f)\mathbf{v}_j$ to the following form:

$$P(f)\mathbf{v}_j = \prod_{i \neq j}^s (h_{i,j})^{n_i} (h_{j,j})^{n_j} \mathbf{v}_j. \quad (7.8)$$

The operator $h_{j,j}$ is a nilpotent operator in the subspace $V(\lambda_j)$ and $n_j = \dim V(\lambda_j)$. Therefore, we can apply the theorem 5.7. As a result we obtain $(h_{j,j})^{n_j} \mathbf{v}_j = \mathbf{0}$. Now from (7.7) and (7.8) for an arbitrary vector $\mathbf{v} \in V$ we derive $P(f)\mathbf{v} = \mathbf{0}$. This proves the theorem for the special case, where the characteristic polynomial of an operator f factorizes into a product of linear terms. The general case is reduced to this special case by means of the field extension technique, which we do not consider in this book. $\square$

CHAPTER III

DUAL SPACE.

§ 1. Linear functionals. Vectors and covectors. Dual space.

DEFINITION 1.1. Let V be a linear vector space over a numeric field $\mathbb{K}$. A numeric function $y = f(\mathbf{v})$ with vectorial argument $\mathbf{v} \in V$ and with values $y \in \mathbb{K}$ is called a *linear functional* if

- (1) $f(\mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{v}_1) + f(\mathbf{v}_2)$ for any two $\mathbf{v}_1, \mathbf{v}_2 \in V$;
- (2) $f(\alpha \cdot \mathbf{v}) = \alpha f(\mathbf{v})$ for any $\mathbf{v} \in V$ and for any $\alpha \in \mathbb{K}$.

The definition of a linear functional is quite similar to the definition of a linear mapping (see definition 8.1 in Chapter I). Comparing these two definitions, we see that any linear functional f is a linear mapping $f: V \rightarrow \mathbb{K}$ and, conversely, any such linear mapping is a linear functional. Thereby the numeric field $\mathbb{K}$ is treated as a linear space of the dimension 1 over itself.

Linear functionals, as linear mappings from V to $\mathbb{K}$, constitute the space $\text{Hom}(V, \mathbb{K})$, which is called the *dual space* or the *conjugate space* for the space V . The dual space $\text{Hom}(V, \mathbb{K})$ is denoted by V^* . The space of homomorphisms $\text{Hom}(V, W)$ is usually determined by two spaces V and W . However, the dual space is an exception $V^* = \text{Hom}(V, \mathbb{K})$, it is determined only by V since $\mathbb{K}$ is known whenever V is given (see definition 2.1 in Chapter I).

Thus, $V^* = \text{Hom}(V, \mathbb{K})$ is a linear vector space over the same numeric field $\mathbb{K}$ as V . If V is finite-dimensional, then the dimension of the conjugate space is determined by the theorem 10.4 in Chapter I: $\dim V^* = \dim V$. The structure of a linear vector space in $V^* = \text{Hom}(V, \mathbb{K})$ is determined by two algebraic operations: the operation of pointwise addition and pointwise multiplication by numbers (see definitions 10.1 and 10.2 in Chapter I). However, it would be worth to formulate these two definitions especially for the present case of linear functionals.

DEFINITION 1.2. Let f and g be two linear functionals of V^* . The *sum* of functionals f and g is a functional h whose values are determined by formula $h(\mathbf{v}) = f(\mathbf{v}) + g(\mathbf{v})$ for all $\mathbf{v} \in V$.

DEFINITION 1.3. Let f be a linear functional of V^* . The product of the functional f by a number $\alpha \in \mathbb{K}$ is a functional h whose values are determined by formula $h(\mathbf{v}) = \alpha \cdot f(\mathbf{v})$ for all $\mathbf{v} \in V$.

Let V be a finite-dimensional vector space over a field $\mathbb{K}$ and let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis in V . Then each vector $\mathbf{v} \in V$ can be expanded in this basis:

$$\mathbf{v} = v^1 \cdot \mathbf{e}_1 + \dots + v^n \cdot \mathbf{e}_n. \quad (1.1)$$

Let's consider i -th coordinate of the vector $\mathbf{v}$. Due to the uniqueness of the expansion (1.1), when the basis is fixed, v^i is a number uniquely determined by the vector $\mathbf{v}$. Hence, we can consider a map $h^i: V \rightarrow \mathbb{K}$, defining it by formula $h^i(\mathbf{v}) = v^i$. When adding vectors, their coordinates are added; when multiplying a vector by a number, its coordinates are multiplied by that number (see the relationships (5.4) in Chapter I). Therefore, $h^i: V \rightarrow \mathbb{K}$ is a linear mapping. This means that each basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of a linear vector space V determines n linear functionals in V^* . The functionals $h^1, \dots, h^n$ are called the *coordinate functionals* the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. They satisfy the relationships

$$h^i(\mathbf{e}_j) = \delta_j^i, \quad (1.2)$$

where δ_j^i is the Kronecker symbol. These relationships (1.2) are called the *relationships of biorthogonality*.

The proof of the relationships of biorthogonality is very simple. If we expand the vector $\mathbf{e}_j$ in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$, then its j -th component is equal to unity, while all other components are equal to zero. Note that $h^i(\mathbf{e}_j)$ is a number equal to i -th component of the vector $\mathbf{e}_j$. Therefore, $h^i(\mathbf{e}_j) = 1$ if $i = j$ and $h^i(\mathbf{e}_j) = 0$ in all other cases.

THEOREM 1.1. *Coordinate functionals $h^1, \dots, h^n$ are linearly independent; they form a basis in dual space V^* .*

PROOF. Let's consider a linear combination of the coordinate functionals associated with a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in V and assume that it is equal to zero:

$$\alpha_1 \cdot h^1 + \dots + \alpha_n \cdot h^n = 0. \quad (1.3)$$

Right hand side of (1.3) is zero functional. Its value when applied to the base vector $\mathbf{e}_j$ is equal to zero. Hence, we have

$$\alpha_1 h^1(\mathbf{e}_j) + \dots + \alpha_n h^n(\mathbf{e}_j) = 0. \quad (1.4)$$

Now we use the relationships of biorthogonality (1.2). Due to these relationships among n terms $h^1(\mathbf{e}_j), \dots, h^n(\mathbf{e}_j)$ in left hand side of the equality (1.4) only one term is nonzero: $h_j(\mathbf{e}_j) = 1$. Therefore, (1.4) reduces to $\alpha_j = 0$. But j is an index that runs from 1 to n . Hence, all coefficients of the linear combination (1.3) are zero, i.e. it is trivial and coordinate functionals $h^1, \dots, h^n$ are linearly independent.

In order to complete the proof of the theorem now we could use the equality $\dim V^* = \dim V = n$ and refer to the item (4) of the theorem 4.5 in Chapter I. However, we choose more explicit way and directly prove that coordinate functionals $h^1, \dots, h^n$ span the dual space V^* . Let $f \in V^*$ be an arbitrary linear functional and let $\mathbf{v}$ be an arbitrary vector of V . Then from (1.1) we derive

$$f(\mathbf{v}) = v^1 f(\mathbf{e}_1) + \dots + v^n f(\mathbf{e}_n) = f(e_1) h^1(\mathbf{v}) + \dots + f(e_n) h^n(\mathbf{v}).$$

Here $f(\mathbf{e}_1), \dots, f(\mathbf{e}_n)$ are numeric coefficients from $\mathbb{K}$ and $\mathbf{v}$ is an arbitrary

vector of V . Therefore, the above equality can be rewritten as an equality of linear functionals in the conjugate space V^* :

$$f = f(\mathbf{e}_1) \cdot h^1 + \dots + f(\mathbf{e}_n) \cdot h^n. \quad (1.5)$$

The formula (1.5) shows that an arbitrary function $f \in V^*$ can be represented as a linear combination of coordinate functionals $h^1, \dots, h^n$. Hence, being linearly independent, they form a basis in V^* . The theorem is proved. $\square$

DEFINITION 1.4. The basis $h^1, \dots, h^n$ in V^* formed by coordinate functionals associated with a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in V is called the *dual basis* or the *conjugate basis* for $\mathbf{e}_1, \dots, \mathbf{e}_n$.

DEFINITION 1.5. Let f be a linear functional in a finite-dimensional space V and let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis in this space. The numbers $f_1, \dots, f_n$ determined by the linear functional f according to the formula

$$f_i = f(\mathbf{e}_i) \quad (1.6)$$

are called the *coordinates* or the *components* of f in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$.

As we see in formula (1.5), the numbers (1.6) are the coefficients of the expansion of f in the conjugate basis $h^1, \dots, h^n$. However, in the definition 1.5 they are mentioned as the components of f in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. This is purely terminological trick, it means that we consider $\mathbf{e}_1, \dots, \mathbf{e}_n$ as a primary basis, while the conjugate basis is treated as an auxiliary and complementary thing.

The algebraic operations of addition and multiplication by numbers in the spaces V and V^* are related to each other by the following equalities:

$$\begin{aligned} f(\mathbf{v}_1 + \mathbf{v}_2) &= f(\mathbf{v}_1) + f(\mathbf{v}_2), & f(\alpha \cdot \mathbf{v}) &= \alpha f(\mathbf{v}); \\ (f_1 + f_2)(\mathbf{v}) &= f_1(\mathbf{v}) + f_2(\mathbf{v}), & (\alpha \cdot f)(\mathbf{v}) &= \alpha f(\mathbf{v}). \end{aligned} \quad (1.7)$$

Vectors and linear functionals enter these equalities in a quite similar way. The fact that in the writing $f(\mathbf{v})$ the functional plays the role of a function, while the vector is written as an argument is not so important. Therefore, sometimes the quantity $f(\mathbf{v})$ is denoted differently:

$$f(\mathbf{v}) = \langle f | \mathbf{v} \rangle. \quad (1.8)$$

The writing (1.8) is associated with the special terminology. Functionals from the dual space V^* are called *covectors*, while the expression $\langle f | \mathbf{v} \rangle$ itself is called the *pairing*, or the *contraction*, or even the *scalar product* of a vector and a covector.

The scalar product (1.8) possesses the property of bilinearity: it is linear in its first argument f and in its second argument $\mathbf{v}$. This follows from the relationships (1.7), which are now written as

$$\begin{aligned} \langle f_1 + f_2 | \mathbf{v} \rangle &= \langle f_1 | \mathbf{v} \rangle + \langle f_2 | \mathbf{v} \rangle, & \langle \alpha \cdot f | \mathbf{v} \rangle &= \alpha \langle f | \mathbf{v} \rangle; \\ \langle f | \mathbf{v}_1 + \mathbf{v}_2 \rangle &= \langle f | \mathbf{v}_1 \rangle + \langle f | \mathbf{v}_2 \rangle, & \langle f | \alpha \cdot \mathbf{v} \rangle &= \alpha \langle f | \mathbf{v} \rangle. \end{aligned} \quad (1.9)$$

We have already dealt with the concept of bilinearity earlier in this book (see theorem 1.1 in Chapter II).

The properties (1.9) of the scalar properties (1.8) are analogous to the properties of the scalar product of geometric vectors — it is usually studied in the course analytic geometry (see [5]). However, in contrast to that «geometric» scalar product, the scalar product (1.8) is not symmetric: its arguments belong to different spaces — they cannot be swapped. Covectors in the scalar product (1.8) are always written on the left and vectors are always on the right.

The following definition is dictated by the intension to strengthen the analogy of (1.8) and traditional «geometric» scalar product.

DEFINITION 1.7. A vector $\mathbf{v}$ and a covector f are called *orthogonal* to each other if their scalar product is zero: $\langle f | \mathbf{v} \rangle = 0$.

THEOREM 1.2. Let $U \subsetneq V$ be a subspace in a finite-dimensional vector space V and let $\mathbf{v} \notin U$. Then there exists a linear functional f in V^* such that $f(\mathbf{v}) \neq 0$ and $f(\mathbf{u}) = 0$ for all $\mathbf{u} \in U$.

PROOF. Let $\dim V = n$ and $\dim U = s$. Let's choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ in a subspace U . Let's add the vector $\mathbf{v}$ to basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_s$ and denote it $\mathbf{v} = \mathbf{e}_{s+1}$. The extended system of vectors is linearly independent since $\mathbf{v} \notin U$, see the item (4) of the theorem 3.1 in Chapter I. Denote by $W = \langle \mathbf{e}_1, \dots, \mathbf{e}_{s+1} \rangle$ the linear span of this system of vectors. It is clear that W is a subspace of V comprising the initial subspace U ; its dimension is one as greater than the dimension of U . The vectors $\mathbf{e}_1, \dots, \mathbf{e}_{s+1}$ form a basis in W . If $W \neq V$, then we complete the basis $\mathbf{e}_1, \dots, \mathbf{e}_{s+1}$ up to a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in the space V and consider the coordinate functionals $h^1, \dots, h^n$ associated with this base. Let's denote $f = h^{s+1}$. Then from the relationships of biorthogonality (1.2) we derive

$$f(\mathbf{v}) = h^{s+1}(\mathbf{e}_{s+1}) = 1 \quad \text{and} \quad f(\mathbf{e}_i) = 0 \quad \text{for } i = 1, \dots, s.$$

Being zero on the basis vectors of the subspace U , the functional $f = h^{s+1}$ vanishes on all vector $\mathbf{u} \in U$. Its value on the vector $\mathbf{v}$ is equal to unity. $\square$

Let's consider the case $U = \{\mathbf{0}\}$ in the above theorem. Then for any nonzero vector $\mathbf{v}$ we have $\mathbf{v} \notin U$, and we can formulate the following corollary of the theorem 1.2.

COROLLARY. For any vector $\mathbf{v} \neq \mathbf{0}$ in a finite-dimensional space V there is a linear functional f in V^* such that $f(\mathbf{v}) \neq 0$.

Let V be a linear vector space over the field $\mathbb{K}$ and let $W = V^*$ be the conjugate space of V . We know that W is also a linear vector space over the field $\mathbb{K}$. Therefore, it possesses its own conjugate space W^* . With respect to V this is the double conjugate space V^{**} . We can also consider triple conjugate, fourth conjugate, etc. Thus we would have the infinite sequence of conjugate spaces. However, soon we shall see, that in the case of finite-dimensional spaces there is no need to consider the multiple conjugate spaces.

Let $\mathbf{v} \in V$. To any $f \in V^*$ we associate the number $f(\mathbf{v}) \in \mathbb{K}$. Thus we define a mapping $\varphi_{\mathbf{v}}: V^* \rightarrow \mathbb{K}$, which is linear due to the following relationships:

$$\varphi_{\mathbf{v}}(f_1 + f_2) = (f_1 + f_2)(\mathbf{v}) = f_1(\mathbf{v}) + f_2(\mathbf{v}) = \varphi_{\mathbf{v}}(f_1) + \varphi_{\mathbf{v}}(f_2),$$

$$\varphi_{\mathbf{v}}(\alpha \cdot f) = (\alpha \cdot f)(\mathbf{v}) = \alpha f(\mathbf{v}) = \alpha \varphi_{\mathbf{v}}(f).$$

Hence, $\varphi_{\mathbf{v}}$ is a linear functional in the space V^* or, in other words, it is an element of double conjugate space. The functional $\varphi_{\mathbf{v}}$ is determined by a vector $\mathbf{v} \in V$. Therefore, when associating $\varphi_{\mathbf{v}}$ with a vector $\mathbf{v}$, we define a mapping

$$h : V \rightarrow V^{**}, \text{ where } h(\mathbf{v}) = \varphi_{\mathbf{v}} \text{ for all } \mathbf{v} \in V. \quad (1.10)$$

The mapping (1.10) is a linear mapping. In order to prove this fact we should verify the following identities for this mapping h :

$$h(\mathbf{v}_1 + \mathbf{v}_2) = h(\mathbf{v}_1) + h(\mathbf{v}_2), \quad h(\alpha \cdot \mathbf{v}) = \alpha h(\mathbf{v}). \quad (1.11)$$

The result of applying h to a vector of the space V is an element of double conjugate space V^{**} . Therefore, in order to verify the equalities (1.11) we should apply both sides of these equalities to an arbitrary covector $f \in V^*$ and check the coincidence of the results that we obtain:

$$\begin{aligned} h(\mathbf{v}_1 + \mathbf{v}_2)(f) &= \varphi_{\mathbf{v}_1 + \mathbf{v}_2}(f) = f(\mathbf{v}_1) + f(\mathbf{v}_2) = \varphi_{\mathbf{v}_1}(f) + \\ &+ \varphi_{\mathbf{v}_2}(f) = h(\mathbf{v}_1)(f) + h(\mathbf{v}_2)(f) = (h(\mathbf{v}_1) + h(\mathbf{v}_2))(f), \\ h(\alpha \cdot \mathbf{v})(f) &= \varphi_{\alpha \cdot \mathbf{v}}(f) = f(\alpha \cdot \mathbf{v}) = \alpha f(\mathbf{v}) = \\ &= \alpha \varphi_{\mathbf{v}}(f) = \alpha h(\mathbf{v})(f) = (\alpha \cdot h(\mathbf{v}))(f). \end{aligned}$$

THEOREM 1.3. *For a finite-dimensional linear vector space V the mapping (1.10) is bijective. It is an isomorphism of the spaces V and V^{**} . This isomorphism is called *canonic isomorphism* of these spaces.*

PROOF. First of all we shall prove the injectivity of the mapping (1.10). For this purpose we consider its kernel $\text{Ker } h$. Let $\mathbf{v}$ be an arbitrary vector of $\text{Ker } h$. Then $\varphi_{\mathbf{v}} = h(\mathbf{v}) = 0$. But $\varphi_{\mathbf{v}} \in V^{**}$, this means that φ is a linear functional in the space V^* . Therefore, the equality $\varphi_{\mathbf{v}} = 0$ means that $\varphi(f) = 0$ for any covector $f \in V^*$. Using this equality, from (1.10) we derive

$$h(\mathbf{v})(f) = \varphi_{\mathbf{v}}(f) = f(\mathbf{v}) = 0 \text{ for all } f \in V^*. \quad (1.12)$$

Now let's apply the corollary of the theorem 1.2. If the vector $\mathbf{v}$ would be nonzero, then there would be a functional f such that $f(\mathbf{v}) \neq 0$. This would contradict the above condition (1.12). Hence, $\mathbf{v} = 0$ by contradiction. This means that $\text{Ker } h = \{0\}$ and h is an injective mapping.

In order to prove the surjectivity of the mapping (1.10) we use the theorem 9.4 from Chapter I. According to this theorem

$$\dim(\text{Ker } h) + \dim(\text{Im } h) = \dim V.$$

We have already proved that $\dim(\text{Ker } h) = 0$. Hence, $\dim(\text{Im } h) = \dim V$. Since $\text{Im } h$ is a subspace of V^{**} and $\dim V^{**} = \dim V^* = \dim V$, we have $\text{Im } h = V^{**}$ (see item (3) of theorem 4.5 in Chapter I). This completes the proof of surjectivity of the mapping h and the proof of the theorem in whole. $\square$

Canonic isomorphism (1.10) possesses the property that for any vector $\mathbf{v} \in V$ and for any covector $f \in V^*$ the following equality holds:

$$\langle h(v) | f \rangle = \langle f | v \rangle. \quad (1.13)$$

The equality (1.13) is derived from the definition of h . Indeed, $\langle h(\mathbf{v}) | f \rangle = h(\mathbf{v})(f) = \varphi_{\mathbf{v}}(f) = f(\mathbf{v}) = \langle f | \mathbf{v} \rangle$. The relationship (1.13) distinguishes canonic isomorphism among all other isomorphisms relating the spaces V and V^{**} .

§ 2. Transformation of the coordinates of a covector under a change of basis.

Let V be a finite-dimensional linear vector space and let V^* be the associated dual space. If we treat V^* separately forgetting its relation to V , then a choice of basis and a change of basis in V^* are quite the same as in any other linear vector space. However, the conjugate space V^* is practically never considered separately. The theory of this space should be understood as an extension of the theory of initial space V .

Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis in a linear vector space V . Each such basis of V has the associated basis of coordinate functionals in V^* . Choosing another basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ in V we immediately get another conjugate basis $\tilde{h}^1, \dots, \tilde{h}^n$ in V^* . Let S be the transition matrix for passing from the old basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ to the new basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$. Similarly, denote by P the transition matrix for passing from the old dual basis $h^1, \dots, h^n$ to the new dual basis $\tilde{h}^1, \dots, \tilde{h}^n$. The components of these two transition matrices S and P are used to expand the vectors of «wavy» bases in corresponding «non-wavy» bases:

$$\tilde{e}_j = \sum_{i=1}^n S_j^i \cdot e_i, \quad \tilde{h}^r = \sum_{s=1}^n P_s^r \cdot h^s. \quad (2.1)$$

Note that the second formula (2.1) differs from the standard given by formula (5.5) in Chapter I: the vectors of dual bases in (2.1) are specified by upper indices despite to the usual convention of enumerating the basis vectors. The reason is that the dual space V^* and the dual bases are treated as complementary objects with respect to the initial space V and its bases. We have already seen such deviations from the standard notations in constructing the basis vectors E_j^i in $\text{Hom}(V, W)$ (see proof of the theorem 10.4 in Chapter I).

In spite of the breaking the standard rules in indexing the basis vectors, the formula (2.1) does not break the rules of tensorial notation: the free index r is in the same upper position in both sides of the equality, the summation index s enters twice — once as an upper index and for the second time as a lower index.

THEOREM 2.1. *The transition matrix P for passing from the old conjugate basis $h^1, \dots, h^n$ to the new conjugate basis $\tilde{h}^1, \dots, \tilde{h}^n$ is inverse to the transition matrix S that is used for passing from the old basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ to the new basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$.*

PROOF. In order to prove this theorem we use the biorthogonality relationships (1.2). Substituting (2.1) into these relationships, we get

$$\delta_j^r = h^r(e_j) = \sum_{s=1}^n \sum_{i=1}^n P_s^r S_j^i h^s(e_i) = \sum_{s=1}^n \sum_{i=1}^n P_s^r S_j^i \delta_i^s = \sum_{i=1}^n P_i^r S_j^i.$$

The above relationship can be written in matrix form $PS = 1$. This means that $P = S^{-1}$. The theorem is proved. $\square$

Remember that the inverse transition matrix T is also the inverse matrix for S . Therefore, in order to write the complete set of formulas relating two pairs of bases in V and V^* it is sufficient to know two matrices S and $T = S^{-1}$:

$$\begin{aligned} \tilde{\mathbf{e}}_j &= \sum_{i=1}^n S_j^i \cdot \mathbf{e}_i, & \tilde{h}^r &= \sum_{s=1}^n T_s^r \cdot h^s, \\ \mathbf{e}_i &= \sum_{j=1}^n T_i^j \cdot \tilde{\mathbf{e}}_j, & h^s &= \sum_{r=1}^n S_r^s \cdot \tilde{h}^r. \end{aligned} \quad (2.2)$$

Let f be a covector from the conjugate space V^* . Let's consider its expansions in two conjugate bases $h^1, \dots, h^n$ and $\tilde{h}^1, \dots, \tilde{h}^n$:

$$f = \sum_{s=1}^n f_s \cdot h^s, \quad f = \sum_{r=1}^n \tilde{f}_r \cdot \tilde{h}^r. \quad (2.3)$$

The expansions (2.3) also differ from the standard introduced by formula (5.1) in Chapter I. To the coordinates of covectors the other standard is applied: they are specified by lower indices and are written in row vectors.

THEOREM 2.2. *The coordinates of a covector f in two dual bases $h^1, \dots, h^n$ and $\tilde{h}^1, \dots, \tilde{h}^n$ associated with the bases $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ in V are related to each other by formulas*

$$\tilde{f}_r = \sum_{s=1}^n S_r^s f_s, \quad f_s = \sum_{j=1}^n T_s^r \tilde{f}_r, \quad (2.4)$$

where S is the direct transition matrix for passing from $\mathbf{e}_1, \dots, \mathbf{e}_n$ to the «wavy» basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$, while $T = S^{-1}$ is the inverse transition matrix.

PROOF. In order to prove the first relationship (2.4) we substitute the fourth expression (2.2) for h^s into the first expansion (2.3):

$$f = \sum_{s=1}^n f_s \cdot \left(\sum_{r=1}^n S_r^s \cdot \tilde{h}^r \right) = \sum_{r=1}^n \left(\sum_{s=1}^n S_r^s f_s \right) \cdot \tilde{h}^r.$$

Then we compare the resulting expansion of f with the second expansion (2.3) and derive the first formula (2.4). The second formula (2.4) is derived similarly. $\square$

Note that the formulas (2.4) can be derived immediately from the definition 1.5 and from formula (1.6) without using the conjugate bases.

THEOREM 2.3. *The scalar product of a vector $\mathbf{v}$ and a covector f is determined by their coordinates according to the formula*

$$\langle f | \mathbf{v} \rangle = \sum_{i=1}^n f_i v^i = f_1 v^1 + \dots + f_n v^n. \quad (2.5)$$

PROOF. In order to prove (2.5) we use the relationship (1.6):

$$\langle f | \mathbf{v} \rangle = f(\mathbf{v}) = \sum_{i=1}^n f(\mathbf{e}_i) v^i = \sum_{i=1}^n f_i v^i.$$

In (2.5) and in the above calculations f is assumed to be expanded in the basis $h^1, \dots, h^n$ conjugated to the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$, where $\mathbf{v}$ is expanded. $\square$

§ 3. Orthogonal complements in a dual space.

DEFINITION 3.1. Let S be a subset in a linear vector space V . The *orthogonal complement* of the subset S in the conjugate space V^* is the set $S^\perp \subset V^*$ composed by covectors each of which orthogonal to all vectors of S .

The above definition of the orthogonal complement $S^\perp$ can be expressed by the formula $S^\perp = \{f \in V^* : \forall \mathbf{v} ((\mathbf{v} \in S) \Rightarrow (\langle f | \mathbf{v} \rangle = 0))\}$.

THEOREM 3.1. The operation of constructing orthogonal complements of subsets $S \subset V$ in the conjugate space V^* possesses the following properties:

- (1) $S^\perp$ is a subspace in V^* ;
- (2) $S_1 \subset S_2$ implies $(S_2)^\perp \subset (S_1)^\perp$;
- (3) $\langle S \rangle^\perp = S^\perp$, where $\langle S \rangle$ is the linear span of S ;
- (4) $\left(\bigcup_{i \in I} S_i \right)^\perp = \bigcap_{i \in I} (S_i)^\perp$.

PROOF. Let's prove the first item of the theorem for the beginning. For this purpose we should verify two conditions from the definition of a subspace. Let $f_1, f_2 \in S^\perp$, then $\langle f_1 | \mathbf{v} \rangle = 0$ and $\langle f_2 | \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in S$. Therefore, for all vectors $\mathbf{v} \in S$ we derive the equality $\langle f_1 + f_2 | \mathbf{v} \rangle = \langle f_1 | \mathbf{v} \rangle + \langle f_2 | \mathbf{v} \rangle = 0$ which means that $f_1 + f_2 \in S^\perp$.

Now assume that $f \in S^\perp$. Then $\langle f | \mathbf{v} \rangle = 0$ for all vectors $\mathbf{v} \in S$. Hence, for the covector $\alpha \cdot f$ we define $\langle \alpha \cdot f | \mathbf{v} \rangle = \alpha \langle f | \mathbf{v} \rangle = 0$. This means that $\alpha \cdot f \in S^\perp$. Thus, the first item in the theorem 3.1 is proved.

In order to prove the inclusion $(S_2)^\perp \subset (S_1)^\perp$ in the second item of the theorem 3.1 we consider an arbitrary covector f of $(S_2)^\perp$. From the condition $f \in (S_2)^\perp$ we derive $\langle f | \mathbf{v} \rangle = 0$ for any $\mathbf{v} \in S_2$. But $S_1 \subset S_2$, therefore, the equality $\langle f | \mathbf{v} \rangle = 0$ holds for any $\mathbf{v} \in S_1$. Then $f \in (S_1)^\perp$. This means that $f \in (S_2)^\perp$ implies $f \in (S_1)^\perp$. The required inclusion is proved.

In order to prove the third item of the theorem note that the linear span of S comprises this set: $S \subset \langle S \rangle$. Applying the item (2) of the theorem, which we have already proved, we obtain the inclusion $\langle S \rangle^\perp \subset S^\perp$. Now we need the opposite inclusion $S^\perp \subset \langle S \rangle^\perp$. In order to prove it let's remember that the linear span $\langle S \rangle$ consists of all possible linear combinations of the form

$$\mathbf{v} = \alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_r \cdot \mathbf{v}_r, \text{ where } \mathbf{v}_i \in S. \quad (3.1)$$

Let $f \in S^\perp$, then $\langle f | \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in S$. In particular, this applies to the vectors $\mathbf{v}_i$ in the expansion (3.1), i.e. $\langle f | \mathbf{v}_i \rangle = 0$. Then from (3.1) we derive

$$\langle f | \mathbf{v} \rangle = \alpha_1 \langle f | \mathbf{v}_1 \rangle + \dots + \alpha_r \langle f | \mathbf{v}_r \rangle = 0.$$

This means that $\langle f | \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in \langle S \rangle$. This proves the opposite inclusion $S^\perp \subset \langle S \rangle^\perp$ and, thus, completes the proof of the equality $\langle S \rangle^\perp = S^\perp$.

Now let's proceed to the proof of the fourth item of the theorem 3.1. For this purpose we introduce the following notations:

$$S = \bigcup_{i \in I} S_i, \quad \tilde{S} = \bigcap_{i \in I} (S_i)^\perp.$$

Let $f \in S^\perp$. Then $\langle f | \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in S$. But $S_i \subset S$ for any $i \in I$. Therefore, $\langle f | \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in S_i$ and for all $i \in I$. This means that f belongs to each of the orthogonal complement $(S_i)^\perp$, therefore, it belongs to their intersection. Thus, we have proved the inclusion $S^\perp \subset \tilde{S}$.

Conversely, from the inclusion $f \in (S_i)^\perp$ for all $i \in I$ we derive $\langle f | \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in S_i$ and for all $i \in I$. This means that the equality $\langle f | \mathbf{v} \rangle = 0$ holds for all vectors $\mathbf{v}$ in the union of all sets S_i . This proves the converse inclusion $\tilde{S} \subset S^\perp$. Thus, we have proved that $S^\perp = \tilde{S}$. The theorem is proved. $\square$

DEFINITION 3.2. Let S be a subset of conjugate space V^* . The *orthogonal complement* of S in V is the set $S^\perp \in V$ formed by vectors each of which is orthogonal to all covectors of the set S .

The above definition of the orthogonal complement $S^\perp \subset V$ can be expressed by the formula $S^\perp = \{\mathbf{v} \in V : \forall f ((f \in S) \Rightarrow (\langle f | \mathbf{v} \rangle = 0))\}$. For this orthogonal complement one can formulate a theorem quite similar to the theorem 3.1.

THEOREM 3.2. *The operation of constructing orthogonal complements of subsets $S \subset V^*$ in V possesses the following four properties:*

- (1) $S^\perp$ is a subspace in V ;
- (2) $S_1 \subset S_2$ implies $(S_2)^\perp \subset (S_1)^\perp$;
- (3) $\langle S \rangle^\perp = S^\perp$, where $\langle S \rangle$ is a linear span of S ;

$$(4) \left(\bigcup_{i \in I} S_i \right)^\perp = \bigcap_{i \in I} (S_i)^\perp.$$

The proof of this theorem almost literally coincides with the proof of the theorem 3.1. Therefore, here we omit this proof.

THEOREM 3.3. *Let V be a finite-dimensional vector space and suppose that we have a subspaces $U \subset V$ and a subspace $W \subset V^*$. The condition $W = U^\perp$ in the sense of definition 3.1 is equivalent to the condition $U = W^\perp$ in the sense of definition 3.2.*

PROOF. Suppose that $W = U^\perp$ in the sense of definition 3.1. Then for any $w \in W$ and for any $\mathbf{u} \in U$ we have the orthogonality $\langle w | \mathbf{u} \rangle = 0$. By definition $W^\perp$ is the set of all vectors $\mathbf{v} \in V$ such that $\langle w | \mathbf{v} \rangle = 0$ for all covectors $w \in W$. Hence, $\mathbf{u} \in U$ implies $\mathbf{u} \in W^\perp$ and we have the inclusion $U \subset W^\perp$.

However, we need to prove the coincidence $U = W^\perp$. Let's do it by contradiction. Suppose that $U \neq W^\perp$. Then there is a vector $\mathbf{v}_0$ such that $\mathbf{v}_0 \in W^\perp$ and $\mathbf{v}_0 \notin U$. In this case we can apply the theorem 1.2 which says that there is a linear functional f such that it vanishes on all vectors $\mathbf{u} \in U$ and is nonzero on the vector $\mathbf{v}_0$. Then $f \in W$ and $\langle f | \mathbf{v}_0 \rangle \neq 0$, so we have the contradiction to the

condition $\mathbf{v}_0 \in W^\perp$. This contradiction proves that $U = W^\perp$. As a result we have proved that $W = U^\perp$ implies $U = W^\perp$.

Now, conversely, let $U = W^\perp$. Then for any $w \in W$ and for any $\mathbf{u} \in U$ we have the orthogonality $\langle w | \mathbf{u} \rangle = 0$. By definition $U^\perp$ is the set of all covectors f perpendicular to all vectors $\mathbf{u} \in U$. Hence, $w \in W$ implies $w \in U^\perp$ and we have the inclusion $W \subset U^\perp$.

Next step is to prove the coincidence $W = U^\perp$. We shall do it again by contradiction. Assume that $W \neq U^\perp$. Then there is a covector $f_0 \in U^\perp$ such that $f_0 \notin W$. Let's apply the theorem 1.2. In this case it says that there is a linear functional φ in V^{**} such that it vanishes on W and is nonzero on the covector f_0 . Remember that we have the canonic isomorphism $h: V \rightarrow V^{**}$. We apply h^{-1} to φ and get the vector $\mathbf{v} = h^{-1}(\varphi)$. Then we take into account (1.13) which yields $\mathbf{v} \in U$ and $\langle f_0 | \mathbf{v} \rangle \neq 0$. This contradicts to the condition $f_0 \in U^\perp$. Hence, by contradiction, $U = W^\perp$ and $U = W^\perp$ implies $W = U^\perp$. The theorem is completely proved. $\square$

The proposition of the theorem 3.3 can be reformulated as follows: in the case of a finite-dimensional space V for any subspace $U \in V$ and for any subspace $W \in V^*$ the following relationships are valid:

$$(U^\perp)^\perp = U, \quad (W^\perp)^\perp = W. \quad (3.2)$$

For arbitrary subsets $S \in V$ and $R \in V^*$ (not subspaces) in the case of a finite-dimensional space V we have the relationships

$$(S^\perp)^\perp = \langle S \rangle, \quad (R^\perp)^\perp = \langle R \rangle. \quad (3.3)$$

These relationships (3.3) are derived from (3.2) by using the item (3) in theorems 3.1 and 3.2.

THEOREM 3.4. *In the case of a finite-dimensional linear vector space V if U is a subspace of V or if U is a subspace of V^* , then $\dim U + \dim U^\perp = \dim V$.*

PROOF. Due to the relationships (3.2) the second case $U \subset V^*$ in the theorem 3.4 is reduced to the first case $U \subset V$ if we replace U by $U^\perp$. Therefore, we consider only the first case $U \subset V$.

Let $\dim V = n$ and $\dim U = s$. We choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ in the subspace U and complete it up to a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in the subspace V . The basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ determines the conjugate basis $h^1, \dots, h^n$ in V^* . If we specify vectors by their coordinates in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ and if we specify covectors by their coordinates in dual basis $h^1, \dots, h^n$, then we can apply the formula (2.5).

By construction of the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ the subspace U consists of vectors the initial s coordinates of which are deliberate, while the remaining $n - s$ coordinates are equal to zero. Therefore the condition $f \in U^\perp$ means that the equality

$$\langle f | v \rangle = \sum_{i=1}^s f_i v^i = 0$$

should be fulfilled identically for any numbers $v^1, \dots, v^s$. This is the case if and only if the first s coordinates of the covector f are zero. Other $n - s$ coordinates

of f are deliberate. This means that the subspace $U^\perp$ is the linear span of the last $n - s$ basis vectors of the conjugate basis:

$$U^\perp = \langle h^{s+1}, \dots, h^n \rangle.$$

For the dimension of the subspace $U^\perp$ this yields $\dim U^\perp = n - s$, hence, we have the required identity $\dim U + \dim U^\perp = \dim V$. The theorem is proved $\square$

The theorem 3.4 is known as the theorem *on the dimension of orthogonal complements*. As an immediate consequence of this theorem we get

$$\begin{aligned} \{0\}^\perp &= V, & V^\perp &= \{0\}, \\ \{0\}^\perp &= V^*, & (V^*)^\perp &= \{0\}. \end{aligned} \quad (3.4)$$

All these equalities have the transparent interpretation. The first three of the equalities (3.4) can be proved immediately without using the finite-dimensionality of V . The proof of the last equality (3.4) uses the corollary of the theorem 1.2, while this theorem assumes V to be a finite-dimensional space.

THEOREM 3.5. *In the case of a finite-dimensional space V for any family of subspaces in V or in V^* the following relationships are fulfilled*

$$\left(\sum_{i \in I} U_i \right)^\perp = \bigcap_{i \in I} (U_i)^\perp, \quad \left(\bigcap_{i \in I} U_i \right)^\perp = \sum_{i \in I} (U_i)^\perp. \quad (3.5)$$

PROOF. The sum of subspaces is the span of their union. Therefore, the first relationship (3.5) is an immediate consequence of the items (3) and (4) in the theorems 3.1 and 3.2. The finite-dimensionality of V here is not used.

The second relationship (3.5) follows from the first one upon substituting U_i by $(U_i)^\perp$. Indeed, applying (3.2), we derive the equality

$$\left(\sum_{i \in I} (U_i)^\perp \right)^\perp = \bigcap_{i \in I} ((U_i)^\perp)^\perp = \bigcap_{i \in I} U_i.$$

Now it is sufficient to pass to orthogonal complements in both sides of this equality and apply (3.2) again. The theorem is proved. $\square$

§ 4. Conjugate mapping.

DEFINITION 4.1. Let $f: V \rightarrow W$ be a linear mapping from V to W . A linear mapping $\varphi: W^* \rightarrow V^*$ is called a *conjugate mapping* for f if for any $\mathbf{v} \in V$ and for any $w \in W^*$ the relationship $\langle \varphi(w) | \mathbf{v} \rangle = \langle w | f(\mathbf{v}) \rangle$ is fulfilled.

The problem of the existence of a conjugate mapping is solved by the definition 4.1 itself. Indeed, in order to define a mapping $\varphi: W^* \rightarrow V^*$ for each functional $w \in W^*$ we should specify the corresponding functional $h = \varphi(w) \in V^*$. But to specify a functional in V^* this means that we should specify its action upon an arbitrary vector $\mathbf{v} \in V$. In the sense of this reasoning the defining relationship for a conjugate mapping is written as follows:

$$h(\mathbf{v}) = \langle h | \mathbf{v} \rangle = \langle \varphi(w) | \mathbf{v} \rangle = \langle w | f(\mathbf{v}) \rangle.$$

It is easy to verify that the above equality defines a linear functional $h = h(\mathbf{v})$:

$$\begin{aligned} h(\mathbf{v}_1 + \mathbf{v}_2) &= \langle w | f(\mathbf{v}_1 + \mathbf{v}_2) \rangle = \langle w | f(\mathbf{v}_1) + f(\mathbf{v}_2) \rangle = \\ &= \langle w | f(\mathbf{v}_1) \rangle + \langle w | f(\mathbf{v}_2) \rangle = h(\mathbf{v}_1) + h(\mathbf{v}_2), \\ h(\alpha \cdot \mathbf{v}) &= \langle w | f(\alpha \cdot \mathbf{v}) \rangle = \langle w | \alpha \cdot f(\mathbf{v}) \rangle = \alpha \langle w | f(\mathbf{v}) \rangle = \alpha h(\mathbf{v}). \end{aligned}$$

THEOREM 4.1. *For a linear mapping $f: V \rightarrow W$ from V to W the conjugate mapping $\varphi: W^* \rightarrow V^*$ is also linear.*

PROOF. Due to the relationship 4.1 for the conjugate mapping $\varphi: W^* \rightarrow V^*$ we have the following relationships:

$$\begin{aligned} \varphi(w_1 + w_2)(\mathbf{v}) &= \langle w_1 + w_2 | f(\mathbf{v}) \rangle = \langle w_1 | f(\mathbf{v}) \rangle + \\ &+ \langle w_2 | f(\mathbf{v}) \rangle = \varphi(w_1)(\mathbf{v}) + \varphi(w_2)(\mathbf{v}) = (\varphi(w_1) + \varphi(w_2))(\mathbf{v}), \\ \varphi(\alpha \cdot w)(\mathbf{v}) &= \langle \alpha \cdot w | f(\mathbf{v}) \rangle = \alpha \langle w | f(\mathbf{v}) \rangle = \\ &= \alpha \varphi(w)(\mathbf{v}) = (\alpha \cdot \varphi(w))(\mathbf{v}). \end{aligned}$$

Since $\mathbf{v} \in V$ is an arbitrary vector of V from the above calculations we obtain $\varphi(w_1 + w_2) = \varphi(w_1) + \varphi(w_2)$ and $\varphi(\alpha \cdot w) = \alpha \cdot \varphi(w)$. This means that the conjugate mapping φ is a linear mapping. $\square$

As we have seen above, the conjugate mapping $\varphi: W^* \rightarrow V^*$ for a mapping $f: V \rightarrow W$ is unique. It is usually denoted $\varphi = f^*$. The operation of passing from f to its conjugate mapping f^* possesses the following properties:

$$(f + g)^* = f^* + g^*, \quad (\alpha \cdot f)^* = \alpha \cdot f^*, \quad (f \circ g)^* = g^* \circ f^*.$$

The first two properties are naturally called the *linearity*. The last third property makes the operation $f \rightarrow f^*$ an analog of the matrix transposition. All three of the above properties are proved by direct calculations on the base of the definition 4.1. We shall not give these calculations here since in what follows we shall not use the above properties at all.

THEOREM 4.2. *In the case of finite-dimensional spaces V and W the kernels and images of the mappings $f: V \rightarrow W$ and $f^*: W^* \rightarrow V^*$ are related as follows:*

$$\begin{aligned} \text{Ker } f^* &= (\text{Im } f)^\perp, & \text{Ker } f &= (\text{Im } f^*)^\perp, \\ \text{Im } f &= (\text{Ker } f^*)^\perp, & \text{Im } f^* &= (\text{Ker } f)^\perp. \end{aligned} \tag{4.1}$$

PROOF. The kernel $\text{Ker } f^*$ is the set of linear functionals of W^* that are mapped to the zero functional in V^* under the action of the mapping f^* . Therefore, $w \in \text{Ker } f^*$ is equivalent to the equality $f^*(w)(\mathbf{v}) = 0$ for all $\mathbf{v} \in V$. As a result of simple calculations we obtain

$$f^*(w)(\mathbf{v}) = \langle f^*(w) | \mathbf{v} \rangle = \langle w | f(\mathbf{v}) \rangle = 0.$$

Hence, the kernel $\text{Ker } f^*$ is the set of covectors orthogonal to the vectors of the form $f(\mathbf{v})$. But the vectors of the form $f(\mathbf{v}) \in W$ constitute the image $\text{Im } f$.

Therefore, $\text{Ker } f^* = (\text{Im } f)^\perp$. The first relationship (4.1) is proved. In proving this relationship we did not use the finite-dimensionality of W . It is valid for infinite dimensional spaces as well.

In order to prove the second relationship we consider the orthogonal complement $(\text{Im } f^*)^\perp$. It is formed by the vectors orthogonal to all covectors of the form $f^*(w)$:

$$0 = \langle f^*(w) | \mathbf{v} \rangle = \langle w | f(\mathbf{v}) \rangle.$$

Using the finite-dimensionality of W , we apply the corollary of the theorem 1.2. It says that if $\langle w | f(\mathbf{v}) \rangle = 0$ for all $w \in W^*$, then $f(\mathbf{v}) = \mathbf{0}$. Therefore, we have $(\text{Im } f^*)^\perp = \text{Ker } f$. The second relationship (4.1) is proved. The third and the fourth relationships are derived from the first and the second ones by means of the theorem 3.3. Thereby we use the finite-dimensionality of the spaces W and V . $\square$

Let the spaces V and W be finite-dimensional. Let's choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in V and a basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_m$ in the space W . This choice uniquely determines the conjugate bases $h^1, \dots, h^n$ and $\tilde{h}^1, \dots, \tilde{h}^m$ in V^* and W^* . Let's consider a mapping $f: V \rightarrow W$ and the conjugate mapping $f^*: W^* \rightarrow V^*$. The matrices of the mappings f and f^* are determined by the expansions:

$$f(\mathbf{e}_j) = \sum_{k=1}^m F_j^k \tilde{\mathbf{e}}_k, \quad f^*(\tilde{h}^i) = \sum_{q=1}^n \Phi_q^i h^q. \quad (4.2)$$

The second relationship (4.1) is somewhat different by structure from the first one. The matter is that the basis vectors of the dual basis are indexed differently (with upper indices). However, this relationship implement the same idea as the first one: the mapping is applied to a basis vector of one space and the result is expanded in the basis of another space.

THEOREM 4.3. *The matrices of the mappings f and f^* determined by the relationships (4.2) are the same, i. e. $F_j^i = \Phi_j^i$.*

PROOF. From the definition of the conjugate mapping we derive

$$\langle \tilde{h}^i | f(\mathbf{e}_j) \rangle = \langle f^*(\tilde{h}^i) | \mathbf{e}_j \rangle. \quad (4.3)$$

Let's calculate separately the left and the right hand sides of this equality using the expansion (4.2) for this purpose:

$$\begin{aligned} \langle \tilde{h}^i | f(\mathbf{e}_j) \rangle &= \sum_{k=1}^m F_j^k \langle \tilde{h}^i | \tilde{\mathbf{e}}_k \rangle = \sum_{k=1}^m F_j^k \delta_k^i = F_j^i, \\ \langle f^*(\tilde{h}^i) | \mathbf{e}_j \rangle &= \sum_{q=1}^n \Phi_q^i \langle h^q | \mathbf{e}_j \rangle = \sum_{q=1}^n \Phi_q^i \delta_j^q = \Phi_j^i. \end{aligned}$$

Substituting the above expressions back to the formula (4.3), we get the required coincidence of the matrices: $F_j^i = \Phi_j^i$. $\square$

Remark. In some theorems of this chapter the restrictions to the finite-dimensional case can be removed. However, the prove of such strengthened versions of these theorems is based on the axiom of choice (see [1]).

CHAPTER IV

BILINEAR AND QUADRATIC FORMS.

§ 1. Symmetric bilinear forms and quadratic forms. Recovery formula.

DEFINITION 1.1. Let V be a linear vector space over a numeric field $\mathbb{K}$. A numeric function $y = f(\mathbf{v}, \mathbf{w})$ with two arguments $\mathbf{v}, \mathbf{w} \in V$ and with the values in the field $\mathbb{K}$ is called a *bilinear form* if

- (1) $f(\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}) = f(\mathbf{v}_1, \mathbf{w}) + f(\mathbf{v}_2, \mathbf{w})$ for any two $\mathbf{v}_1, \mathbf{v}_2 \in V$;
- (2) $f(\alpha \cdot \mathbf{v}, \mathbf{w}) = \alpha f(\mathbf{v}, \mathbf{w})$ for any $\mathbf{v} \in V$ and for any $\alpha \in \mathbb{K}$;
- (3) $f(\mathbf{v}, \mathbf{w}_1 + \mathbf{w}_2) = f(\mathbf{v}, \mathbf{w}_1) + f(\mathbf{v}, \mathbf{w}_2)$ for any two $\mathbf{v}_1, \mathbf{v}_2 \in V$;
- (4) $f(\mathbf{v}, \alpha \cdot \mathbf{w}) = \alpha f(\mathbf{v}, \mathbf{w})$ for any $\mathbf{v} \in V$ and for any $\alpha \in \mathbb{K}$.

The bilinear form $f(\mathbf{v}, \mathbf{w})$ is linear in its first argument $\mathbf{v}$ when the second argument $\mathbf{w}$ is fixed; it is also linear in its second argument $\mathbf{w}$ when the first argument $\mathbf{v}$ is fixed.

DEFINITION 1.2. A bilinear form $f(\mathbf{v}, \mathbf{w})$ is called a *symmetric bilinear form* if $f(\mathbf{v}, \mathbf{w}) = f(\mathbf{w}, \mathbf{v})$.

DEFINITION 1.3. A bilinear form $f(\mathbf{v}, \mathbf{w})$ is called a *skew-symmetric bilinear form* or an *antisymmetric bilinear form* if $f(\mathbf{v}, \mathbf{w}) = -f(\mathbf{w}, \mathbf{v})$.

Having a bilinear form $f(\mathbf{v}, \mathbf{w})$, one can produce a symmetric bilinear form:

$$f_+(\mathbf{v}, \mathbf{w}) = \frac{f(\mathbf{v}, \mathbf{w}) + f(\mathbf{w}, \mathbf{v})}{2}. \quad (1.1)$$

Similarly, one can produce a skew-symmetric bilinear form:

$$f_-(\mathbf{v}, \mathbf{w}) = \frac{f(\mathbf{v}, \mathbf{w}) - f(\mathbf{w}, \mathbf{v})}{2}. \quad (1.2)$$

The operation (1.1) is called the *symmetrization* of the bilinear form f ; the operation (1.2) is called the *alternation* of this bilinear form. Thereby any bilinear form is the sum of a symmetric bilinear form and a skew-symmetric one:

$$f(\mathbf{v}, \mathbf{w}) = f_+(\mathbf{v}, \mathbf{w}) + f_-(\mathbf{v}, \mathbf{w}). \quad (1.3)$$

THEOREM 1.1. The expansion of a given bilinear form $f(\mathbf{v}, \mathbf{w})$ into the sum of a symmetric and a skew-symmetric bilinear forms is unique.

PROOF. Let's consider an expansion of $f(\mathbf{v}, \mathbf{w})$ into the sum of a symmetric and a skew-symmetric bilinear forms

$$f(\mathbf{v}, \mathbf{w}) = h_+(\mathbf{v}, \mathbf{w}) + h_-(\mathbf{v}, \mathbf{w}). \quad (1.4)$$

By means of symmetrization and alternation from (1.4) we derive

$$\begin{aligned} f(v, w) + f(w, v) &= (h_+(v, w) + h_+(w, v)) + \\ &\quad + (h_-(v, w) + h_-(w, v)) = 2h_+(v, w), \\ f(v, w) - f(w, v) &= (h_+(v, w) - h_+(w, v)) + \\ &\quad + (h_-(v, w) - h_-(w, v)) = 2h_-(v, w), \end{aligned}$$

Hence, $h_+ = f_+$ and $h_- = f_-$. Therefore, the expansion (1.4) coincides with the expansion (1.3). The theorem is proved. $\square$

DEFINITION 1.4. A numeric function $y = g(\mathbf{v})$ with one vectorial argument $\mathbf{v} \in V$ is called a *quadratic form* in a linear vector space V if $g(\mathbf{v}) = f(\mathbf{v}, \mathbf{v})$ for some bilinear form $f(\mathbf{v}, \mathbf{w})$.

If $g(\mathbf{v}) = f(\mathbf{v}, \mathbf{v})$, then the quadratic form g is said to be *generated* by the bilinear form f . For a skew-symmetric bilinear form we have $f_-(\mathbf{v}, \mathbf{v}) = -f_-(\mathbf{v}, \mathbf{v})$. Hence, $f_-(\mathbf{v}, \mathbf{v}) = 0$. Then from the expansion (1.3) we derive

$$g(\mathbf{v}) = f(\mathbf{v}, \mathbf{v}) = f_+(\mathbf{v}, \mathbf{v}). \quad (1.5)$$

The same quadratic form can be generated by several bilinear forms. The relationship (1.5) shows that any quadratic form can be generated by a symmetric bilinear form.

THEOREM 1.2. For any quadratic form $g(\mathbf{v})$ there is the unique bilinear form $f(\mathbf{v}, \mathbf{w})$ that generates $g(\mathbf{v})$.

PROOF. The existence of a symmetric bilinear form $f(\mathbf{v}, \mathbf{w})$ generating $g(\mathbf{v})$ follows from (1.5). Let's prove the uniqueness of this form. From $g(\mathbf{v}) = f(\mathbf{v}, \mathbf{v})$ and from the symmetry of the form f we derive

$$\begin{aligned} g(\mathbf{v} + \mathbf{w}) &= f(\mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}) = f(\mathbf{v}, \mathbf{v}) + f(\mathbf{v}, \mathbf{w}) + \\ &\quad + f(\mathbf{w}, \mathbf{v}) + f(\mathbf{w}, \mathbf{w}) = f(\mathbf{v}, \mathbf{v}) + 2f(\mathbf{v}, \mathbf{w}) + f(\mathbf{w}, \mathbf{w}). \end{aligned}$$

Now $f(\mathbf{v}, \mathbf{v})$ and $f(\mathbf{w}, \mathbf{w})$ in right hand side of this formula can be replaced by $g(\mathbf{v})$ and $g(\mathbf{w})$ respectively. Hence, we get

$$f(\mathbf{v}, \mathbf{w}) = \frac{g(\mathbf{v} + \mathbf{w}) - g(\mathbf{v}) - g(\mathbf{w})}{2}. \quad (1.6)$$

Formula (1.6) shows that the values of the symmetric bilinear form $f(\mathbf{v}, \mathbf{w})$ are uniquely determined by the values of the quadratic form $g(\mathbf{v})$. This proves the uniqueness of the form f . $\square$

The formula (1.6) is called a *recovery formula*. Usually, a quadratic form and an associated symmetric bilinear form for it both are denoted by the same symbol: $g(\mathbf{v}) = g(\mathbf{v}, \mathbf{v})$. Moreover, when a quadratic form is given, we assume without stipulations that the associated symmetric bilinear form $g(\mathbf{v}, \mathbf{w})$ is also given.

Let $f(\mathbf{v}, \mathbf{w})$ be a bilinear form in a finite-dimensional linear vector space V and let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis in this space. The numbers f_{ij} determined by formula

$$f_{ij} = f(\mathbf{e}_i, \mathbf{e}_j) \quad (1.7)$$

are called the *coordinates* or the *components* of the form f in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. The numbers (1.7) are written in form of a matrix

$$F = \begin{pmatrix} f_{11} & \dots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{n1} & \dots & f_{nn} \end{pmatrix}, \quad (1.8)$$

which is called the matrix of the bilinear form f in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. For the element f_{ij} in the matrix (1.8) the first index i specifies the row number, the second index j specifies the column number. The matrix of a symmetric bilinear form g is also symmetric: $g_{ij} = g_{ji}$. Further, saying the matrix of a quadratic form $g(\mathbf{v})$, we shall assume the matrix of an associated symmetric bilinear form $g(\mathbf{v}, \mathbf{w})$.

Let $v^1, \dots, v^n$ and $w^1, \dots, w^n$ be coordinates of two vectors $\mathbf{v}$ and $\mathbf{w}$ in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. Then the values $f(\mathbf{v}, \mathbf{w})$ and $g(\mathbf{v})$ of a bilinear form and of a quadratic form respectively are calculated by the following formulas:

$$f(v, w) = \sum_{i=1}^n \sum_{j=1}^n f_{ij} v^i w^j, \quad g(v) = \sum_{i=1}^n \sum_{j=1}^n g_{ij} v^i v^j. \quad (1.9)$$

In the case when g_{ij} is a diagonal matrix, the formula for $g(\mathbf{v})$ contains only the squares of coordinates of a vector $\mathbf{v}$:

$$g(\mathbf{v}) = g_{11}(v^1)^2 + \dots + g_{nn}(v^n)^2. \quad (1.10)$$

This supports the term «quadratic form». Bringing a quadratic form to the form (1.10) by means of choosing proper basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in a linear space V is one of the problems which are solved in the theory of quadratic form.

Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ be two bases in a linear vector space V . Let's denote by S the transition matrix for passing from the first basis to the second one. Denote $T = S^{-1}$. From (1.7) we easily derive the formula relating the components of a bilinear form $f(\mathbf{v}, \mathbf{w})$ these two bases. For this purpose it is sufficient to substitute the relationship (5.8) of Chapter I into the formula (1.7) and use the bilinearity of the form $f(\mathbf{v}, \mathbf{w})$:

$$f_{ij} = f(\mathbf{e}_i, \mathbf{e}_j) = \sum_{k=1}^n \sum_{q=1}^n T_i^k T_j^q f(\tilde{\mathbf{e}}_k, \tilde{\mathbf{e}}_q) = \sum_{k=1}^n \sum_{q=1}^n T_i^k T_j^q \tilde{f}_{kq}.$$

The reverse formula expressing $\tilde{f}_{kq}$ through f_{ij} is derived similarly:

$$f_{ij} = \sum_{k=1}^n \sum_{q=1}^n T_i^k T_j^q \tilde{f}_{kq}, \quad \tilde{f}_{kq} = \sum_{i=1}^n \sum_{j=1}^n S_k^i S_q^j f_{ij}. \quad (1.11)$$

In matrix form these relationships are written as follows:

$$F = T^{\text{tr}} \tilde{F} T, \quad \tilde{F} = S^{\text{tr}} F S. \quad (1.12)$$

Here S^{tr} and T^{tr} are two matrices obtained from S and T by transposition.

§ 2. Orthogonal complements with respect to a quadratic form.

DEFINITION 2.1. Two vectors $\mathbf{v}$ and $\mathbf{w}$ in a linear vector space V are called *orthogonal to each other with respect to the quadratic form g* if $g(\mathbf{v}, \mathbf{w}) = 0$.

DEFINITION 2.2. Let S be a subset of a linear vector space V . The *orthogonal complement* of the subset S with respect to a quadratic form $g(\mathbf{v})$ is the set of vectors each of which is orthogonal to all vectors of S with respect to that quadratic form g . The orthogonal complement of S is denoted $S_{\perp} \subset V$.

The orthogonal complement of a subset S with respect to a quadratic form g can be defined formally: $S_{\perp} = \{\mathbf{v} \in V : \forall \mathbf{w} ((\mathbf{w} \in S) \Rightarrow (g(\mathbf{v}, \mathbf{w}) = 0))\}$. For the orthogonal complements determined by a quadratic form $g(\mathbf{v})$ there is a theorem analogous to theorems 3.1 and 3.2 in Chapter III.

THEOREM 2.1. *The operation of constructing orthogonal complements of subsets $S \subset V$ with respect to a quadratic form g possesses the following properties:*

- (1) $S_{\perp}$ is a subspace in V ;
- (2) $S_1 \subset S_2$ implies $(S_2)_{\perp} \subset (S_1)_{\perp}$;
- (3) $\langle S \rangle_{\perp} = S_{\perp}$, where $\langle S \rangle$ is the linear span of S ;

$$(4) \left(\bigcup_{i \in I} S_i \right)_{\perp} = \bigcap_{i \in I} (S_i)_{\perp}.$$

PROOF. Let's prove the first item in the theorem for the beginning. For this purpose we should verify two conditions from the definition of a subspace.

Let $\mathbf{v}_1, \mathbf{v}_2 \in S_{\perp}$. Then $g(\mathbf{v}_1, \mathbf{w}) = 0$ and $g(\mathbf{v}_2, \mathbf{w}) = 0$ for all $\mathbf{w} \in S$. Hence, for all $\mathbf{w} \in S$ we have $g(\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}) = g(\mathbf{v}_1, \mathbf{w}) + g(\mathbf{v}_2, \mathbf{w}) = 0$. This means that $\mathbf{v}_1 + \mathbf{v}_2 \in S_{\perp}$, so the first condition is verified.

Now let $\mathbf{v} \in S_{\perp}$. Then $g(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w} \in S$. Hence, for the vector $\alpha \cdot \mathbf{v}$ we derive $g(\alpha \cdot \mathbf{v}, \mathbf{w}) = \alpha g(\mathbf{v}, \mathbf{w}) = 0$. This means that $\alpha \cdot \mathbf{v} \in S_{\perp}$. Thus, the first item of the theorem 2.1 is proved.

In order to prove the inclusion $(S_2)_{\perp} \subset (S_1)_{\perp}$ in the second item of the theorem 2.1 we consider an arbitrary vector $\mathbf{v}$ in $(S_2)_{\perp}$. From the condition $\mathbf{v} \in (S_2)_{\perp}$ we get $g(\mathbf{v}, \mathbf{w}) = 0$ for any $\mathbf{w} \in S_2$. But $S_1 \subset S_2$, therefore, the equality $g(\mathbf{v}, \mathbf{w}) = 0$ is fulfilled for any $\mathbf{w} \in S_1$. Then $\mathbf{v} \in (S_1)_{\perp}$. Thus, $\mathbf{v} \in (S_2)_{\perp}$ implies $\mathbf{v} \in (S_1)_{\perp}$. This proves the required inclusion.

Now let's proceed to the third item of the theorem. Note that the linear span of S comprises this set: $S \subset \langle S \rangle$. Applying the second item of the theorem, which is already proved, we get the inclusion $\langle S \rangle_{\perp} \subset S_{\perp}$. In order to prove the coincidence $\langle S \rangle_{\perp} = S_{\perp}$ we have to prove the converse inclusion $S_{\perp} \subset \langle S \rangle_{\perp}$. For this purpose let's remember that the linear span $\langle S \rangle$ is formed by the linear combinations

$$\mathbf{w} = \alpha_1 \cdot \mathbf{w}_1 + \dots + \alpha_r \cdot \mathbf{w}_r, \quad \text{where } \mathbf{w}_i \in S. \quad (2.1)$$

Let $\mathbf{v} \in S_\perp$, then $g(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w} \in S$. In particular, this is true for the vectors $\mathbf{w}_i$ in the expansion (2.1): $g(\mathbf{v}, \mathbf{w}_i) = 0$. Then from (2.1) we derive

$$g(\mathbf{v}, \mathbf{w}) = \alpha_1 g(\mathbf{v}, \mathbf{w}_1) + \dots + \alpha_r g(\mathbf{v}, \mathbf{w}_r) = 0.$$

Hence, $g(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w} \in \langle S \rangle$. This proves the converse inclusion $S_\perp \subset \langle S \rangle_\perp$ and thus completes the proof of the coincidence $\langle S \rangle_\perp = S_\perp$.

In proving the fourth item of the theorem we introduce the following notations:

$$S = \bigcup_{i \in I} S_i, \quad \tilde{S} = \bigcap_{i \in I} (S_i)_\perp.$$

Let $\mathbf{v} \in S_\perp$. Then $g(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w} \in S$. But $S_i \subset S$ for any $i \in I$. Therefore, $g(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w} \in S_i$ and for all $i \in I$. This means that $\mathbf{v}$ belongs to each of the orthogonal complements $(S_i)_\perp$ and, hence, it belongs to their intersection. This proves the inclusion $S_\perp \subset \tilde{S}$.

Conversely, if $\mathbf{v} \in (S_i)_\perp$ for all $i \in I$, then $g(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w} \in S_i$ and for all $i \in I$. Hence, $g(\mathbf{v}, \mathbf{w}) = 0$ for any vector $\mathbf{w}$ in the union of all sets S_i . This proves the converse inclusion $\tilde{S} \subset S_\perp$.

The above two inclusions $S_\perp \subset \tilde{S}$ and $\tilde{S} \subset S_\perp$ prove the coincidence of two sets $S_\perp = \tilde{S}$. The theorem 2.1 is proved. $\square$

DEFINITION 2.3. The *kernel* of a quadratic form $g(\mathbf{v})$ in a linear vector space V is the set $\text{Ker } g = V_\perp$ formed by vectors orthogonal to each vector of the space V with respect to the form g .

DEFINITION 2.4. A quadratic form with nontrivial kernel $\text{Ker } g \neq \{\mathbf{0}\}$ is called a *degenerate* quadratic form. Otherwise, if $\text{Ker } g = \{\mathbf{0}\}$, then the form g is called a *non-degenerate* quadratic form.

Due to the theorem 2.1 the kernel of a form $g(\mathbf{v})$ is a subspace of the space V , where it is defined. The term «kernel» is not an occasional choice for denoting the set $V_\perp$. Each quadratic form is associated with some mapping, for which the subspace $V_\perp$ is the kernel.

DEFINITION 2.5. An *associated mapping* of a quadratic form g is the mapping $a_g: V \rightarrow V^*$ that takes each vector $\mathbf{v}$ of the space V to the linear functional $f_{\mathbf{v}}$ in the conjugate space V^* determined by the relationship

$$f_{\mathbf{v}}(\mathbf{w}) = g(\mathbf{v}, \mathbf{w}) \quad \text{for all } \mathbf{w} \in V. \quad (2.2)$$

The associated mapping $a_g: V \rightarrow V^*$ is linear, this fact is immediate from the bilinearity of the form g . Its kernel $\text{Ker } a_g$ coincides with the kernel of the form g . Indeed, the condition $\mathbf{v} \in \text{Ker } a_g$ means that the functional $f_{\mathbf{v}}$ determined by (2.2) is identically zero. Hence, $\mathbf{v}$ is orthogonal to all vectors $\mathbf{w} \in V$ with respect to the quadratic form $g(\mathbf{v})$.

The associated mapping a_g relates orthogonal complements $S_\perp$ determined by the quadratic form g and orthogonal complements $S^\perp$ in a dual space, which we considered earlier in Chapter III.

THEOREM 2.2. *For any subset $S \subset V$ and for any quadratic form $g(\mathbf{v})$ in a linear vector space V the set $S_\perp$ is the total preimage of the set $S^\perp$ under the associated mapping a_g , i. e. $S_\perp = a_g^{-1}(S^\perp)$.*

PROOF. The condition $\mathbf{v} \in S_\perp$ means that $g(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w} \in S$. But this equality can be rewritten in the following way:

$$g(\mathbf{v}, \mathbf{w}) = f_{\mathbf{v}}(\mathbf{w}) = a_g(\mathbf{v})(\mathbf{w}) = \langle a_g(\mathbf{v}) | \mathbf{w} \rangle = 0 \quad \text{for all } \mathbf{w} \in S.$$

Hence, the condition $\mathbf{v} \in S_\perp$ is equivalent to $a_g(\mathbf{v}) \in S^\perp$. This proves the required equality $S_\perp = a_g^{-1}(S^\perp)$. $\square$

According to the definition 2.3, vectors of the kernel $\text{Ker } g$ are orthogonal to all vectors of the space V with respect to the form g . Therefore $(\text{Ker } g)_\perp = V$. If we apply the result of the theorem 2.2 to the kernel $S = \text{Ker } g$, we get

$$a_g^{-1}((\text{Ker } g)^\perp) = (\text{Ker } g)_\perp = V.$$

This result becomes more clear if we write it in the following equivalent form:

$$\text{Im } a_g = a_g(V) \subseteq (\text{Ker } g)^\perp. \quad (2.3)$$

COROLLARY 1. *The image of the associated mapping a_g is enclosed into the orthogonal complement to its kernel $(\text{Ker } a_g)^\perp$, i. e. $\text{Im } a_g \subseteq (\text{Ker } a_g)^\perp$.*

This corollary of the theorem 2.2 is derived from the formula (2.3) if we take into account $\text{Ker } g = \text{Ker } a_g$. For a quadratic form g in a finite-dimensional space V it can be strengthened.

COROLLARY 2. *For a quadratic form $g(\mathbf{v})$ in a finite-dimensional linear vector space V the image of the associated mapping $a_g : V \rightarrow V^*$ coincides with the orthogonal complement of its kernel $\text{Ker } a_g$:*

$$\text{Im } a_g = (\text{Ker } a_g)^\perp. \quad (2.4)$$

PROOF. Using the theorem 9.4 from Chapter I, we calculate the dimension of the image $\text{Im } a_g$ of the associated mapping:

$$\dim(\text{Im } a_g) = \dim V - \dim(\text{Ker } a_g).$$

The dimension of the orthogonal complement of $\text{Ker } a_g$ in the dual space is determined by the theorem 3.4 in Chapter III:

$$\dim(\text{Ker } a_g)^\perp = \dim V - \dim(\text{Ker } a_g).$$

As we can see, the dimensions of these two subspaces are equal to each other. Therefore, we can apply the above corollary 1 and the item (3) of the theorem 4.5 from Chapter I. As a result we get the required equality (2.4). $\square$

THEOREM 2.3. *Let $U \subsetneq V$ be a subspace of a finite-dimensional space V comprising the kernel of a quadratic form g . For any vector $\mathbf{v} \notin U$ there exists a vector $\mathbf{w} \in V$ such that $g(\mathbf{v}, \mathbf{w}) \neq 0$ and $g(\mathbf{v}, \mathbf{u}) = 0$ for all $\mathbf{u} \in U$.*

PROOF. This theorem is an analog of the theorem 1.2 from Chapter III. It's proof is essentially based on that theorem. Applying the theorem 1.2 from Chapter III, we get that there exist a linear functional $f \in V^*$ such that $f(\mathbf{v}) \neq 0$ and $f(\mathbf{u}) = \langle f | \mathbf{u} \rangle = 0$ for all $\mathbf{u} \in U$. Due to the last condition this functional f belongs to the orthogonal complement $U^\perp$. From the inclusion $\text{Ker } g \subset U$, applying the item (2) of the theorem 3.1 from Chapter III, we get $U^\perp \subset (\text{Ker } g)^\perp$. Hence, we conclude that $f \in (\text{Ker } g)^\perp$.

Now we apply the corollary 2 from the theorem 2.2. From this corollary we obtain that $(\text{Ker } g)^\perp = \text{Im } a_g$. Hence, $f \in \text{Im } a_g$ and there is a vector $\mathbf{w} \in V$ that is taken to f by the associated mapping a_g , i. e. $f = a_g(\mathbf{w})$. Then

$$\begin{aligned} g(\mathbf{v}, \mathbf{w}) &= a_g(\mathbf{w})(\mathbf{v}) = f(\mathbf{v}) \neq 0, \\ g(\mathbf{v}, \mathbf{u}) &= a_g(\mathbf{w})(\mathbf{u}) = f(\mathbf{u}) = 0 \quad \text{for all } \mathbf{u} \in U. \end{aligned}$$

Due to these relationship we find that $\mathbf{w}$ is the very vector that we need to complete the proof of the theorem. $\square$

THEOREM 2.4. *Let V be a finite-dimensional linear vector space and let U and W be two subspaces of V comprising the kernel $\text{Ker } g$ of a quadratic form g . Then the conditions $W = U_\perp$ and $U = W_\perp$ are equivalent to each other.*

PROOF. The theorem 2.4 is an analog of the theorem 3.3 from Chapter III. The proofs of these two theorems are also very similar.

Suppose that the condition $W = U_\perp$ is fulfilled. Then for any vector $\mathbf{w} \in W$ and for any vector $\mathbf{u} \in U$ we have the relationship $g(\mathbf{w}, \mathbf{u}) = 0$. The set $W_\perp$ is formed by vectors orthogonal to all vectors of W with respect to the quadratic form g . Therefore, we have the inclusion $U \subset W_\perp$.

Further proof is by contradiction. Assume that $U \neq W_\perp$. Then there is a vector $\mathbf{v}_0$ such that $\mathbf{v}_0 \in W_\perp$ and $\mathbf{v}_0 \notin U$. In this situation we can apply the theorem 2.3 which says that there exists a vector $\mathbf{v}$ such that $g(\mathbf{v}, \mathbf{v}_0) \neq 0$ and $g(\mathbf{v}, \mathbf{u}) = 0$ for all $\mathbf{u} \in U$. The latter condition means that $\mathbf{v} \in U_\perp = W$. Then the other condition $g(\mathbf{v}, \mathbf{v}_0) \neq 0$ contradicts to the initial choice $\mathbf{v}_0 \in W_\perp$. This contradiction shows that the assumption $U \neq W_\perp$ is not true and we have the coincidence $U = W_\perp$. Thus, $W = U_\perp$ implies $U = W_\perp$. We can swap U and W and obtain that $U = W_\perp$ implies $W = U_\perp$. Hence, these two conditions are equivalent. $\square$

The proposition of the theorem 2.3 can be reformulated as follows: for a subspace $U \subset V$ in a finite-dimensional space V the condition $\text{Ker } g \subset U$ means that double orthogonal complement of U coincides with that space: $(U_\perp)_\perp = U$. For an arbitrary subset $S \subset V$ of a finite-dimensional space V one can derive

$$(S_\perp)_\perp = \langle S \rangle + \text{Ker } g. \quad (2.5)$$

Let's prove the relationship (2.5). Note that vectors of the kernel $\text{Ker } g$ are orthogonal to all vectors of V . Therefore, joining the vectors of the kernel $\text{Ker } g$

to S , we do not change the orthogonal complement of this subset:

$$S_{\perp} = (S \cup \text{Ker } g)_{\perp}.$$

Now let's apply the item (3) of the theorem 2.1. This yields

$$S_{\perp} = (S \cup \text{Ker } g)_{\perp} = \langle S \cup \text{Ker } g \rangle_{\perp} = (\langle S \rangle + \text{Ker } g)_{\perp}.$$

The subspace $U = \langle S \rangle + \text{Ker } g$ comprises the kernel of the form g . Therefore, $(U_{\perp})_{\perp} = U$. This completes the proof of the relationship (2.5):

$$(S_{\perp})_{\perp} = ((\langle S \rangle + \text{Ker } g)_{\perp})_{\perp} = \langle S \rangle + \text{Ker } g.$$

THEOREM 2.5. *In the case of finite-dimensional linear vector space V for any subspace U of V we have the equality*

$$\dim U + \dim U_{\perp} = \dim V + \dim(\text{Ker } g \cap U), \quad (2.6)$$

where $U_{\perp}$ is the orthogonal complement of U with respect to the form g .

PROOF. The vectors of the kernel $\text{Ker } g$ are orthogonal to all vectors of the space V , therefore, joining them to U , we do not change the orthogonal complement $U_{\perp}$. Let's denote $W = U + \text{Ker } g$. Then $U_{\perp} = W_{\perp}$. Applying the theorem 6.4 from Chapter I, for the dimension of W we derive the formula

$$\dim W = \dim U + \dim(\text{Ker } g) - \dim(\text{Ker } g \cap U). \quad (2.7)$$

Now let's apply the theorem 2.2 to the subset $S = W$. This yields $W_{\perp} = a_g^{-1}(W^{\perp})$. Note that $\text{Ker } g \subset W$, this differs W from the initial subspace U . Let's apply the item (2) of the theorem 3.1 to the inclusion $\text{Ker } g \subset W$ and take into account the corollary 2 of the theorem 2.2. This yields

$$W^{\perp} \subset (\text{Ker } g)^{\perp} = \text{Im } a_g.$$

The inclusion $W^{\perp} \subset \text{Im } a_g$ means that the preimage of each element $f \in W^{\perp}$ under the mapping a_g is not empty, while the equality $W_{\perp} = a_g^{-1}(W^{\perp})$ shows that such preimage is enclosed into $W_{\perp}$. Therefore, $W_{\perp} = a_g^{-1}(W^{\perp})$ implies the equality $a_g(W_{\perp}) = W^{\perp}$.

Now let's consider the restriction of the associated mapping a_g to the subspace $W_{\perp}$. We denote this restriction by a :

$$a: W_{\perp} \rightarrow V^*. \quad (2.8)$$

The kernel of the mapping (2.8) coincides with the kernel of the non-restricted mapping a_g since $\text{Ker } a_g = \text{Ker } g \subset W_{\perp}$. For the image of this mapping we have

$$\text{Im } a = a_g(W_{\perp}) = W^{\perp}.$$

Let's apply the theorem on the sum of dimensions of the kernel and the image (see theorem 9.4 in Chapter I) to the mapping a :

$$\dim(\text{Ker } g) + \dim W^{\perp} = \dim W_{\perp} \quad (2.9)$$

In order to determine the dimension of $W^\perp$ we apply the relationship

$$\dim W + \dim W^\perp = \dim V \quad (2.10)$$

which follows from the theorem 3.4 of Chapter III. Now let's add the relationships (2.7) and (2.9) and subtract the relationship (2.10). Taking into account the coincidence $W_\perp = U_\perp$, we get the required equality (2.6). $\square$

The analogs of the relationships (3.4) from Chapter III in present case are the relationships $\{\mathbf{0}\}_\perp = V$ and $V_\perp = \text{Ker } g$.

THEOREM 2.6. *In the case of finite-dimensional linear vector space V equipped with a quadratic form g for any family of subspaces in V , each of which comprises the kernel $\text{Ker } g$, the following relationships are fulfilled:*

$$\left(\sum_{i \in I} U_i \right)_\perp = \bigcap_{i \in I} (U_i)_\perp, \quad \left(\bigcap_{i \in I} U_i \right)_\perp = \sum_{i \in I} (U_i)_\perp. \quad (2.11)$$

PROOF. In proving the first relationship (2.11) the condition $\text{Ker } g \subset U_i$ is inessential. This relationship is derived from the items (3) and (4) of the theorem 2.1 if we take into account that the sum of subspaces is the linear span of the union of these subspaces.

The second relationship (2.11) is derived from the first one. From the condition $\text{Ker } g \in U_i$ we derive that $((U_i)_\perp)_\perp = U_i$ (see theorem 2.4). Let's denote $(U_i)_\perp = V_i$ and apply the first relationship (2.11) to the family of subsets V_i :

$$\left(\sum_{i \in I} (U_i)_\perp \right)_\perp = \left(\sum_{i \in I} V_i \right)_\perp = \bigcap_{i \in I} (V_i)_\perp = \bigcap_{i \in I} U_i.$$

Now it is sufficient to pass to orthogonal complements in left and right hand sides of the above equality and apply the theorem 2.4 again. This yields the required equality (2.11). The theorem is proved. $\square$

§ 3. Transformation of a quadratic form to its canonic form. Inertia indices and signature.

DEFINITION 3.1. A subspace U in a linear vector space V is called *regular with respect to a quadratic form g* if $U \cap U_\perp \subseteq \text{Ker } g$.

THEOREM 3.1. *Let U be a subspace in a finite-dimensional space V regular with respect to a quadratic form g . Then $U + U_\perp = V$.*

PROOF. Let's denote $W = U + U_\perp$ and then let's calculate the dimension of the subspace W applying the theorem 6.4 from Chapter I:

$$\dim W = \dim U + \dim U_\perp - \dim(U \cap U_\perp).$$

The vectors of the kernel $\text{Ker } g$ are perpendicular to all vectors of the space V . Therefore, $\text{Ker } g \subseteq U_\perp$. Moreover, due to the regularity of U with respect to the form g we have $U \cap U_\perp \subseteq \text{Ker } g$. Therefore, we derive

$$U \cap U_\perp = (U \cap U_\perp) \cap \text{Ker } g = U \cap (U_\perp \cap \text{Ker } g) = U \cap \text{Ker } g.$$

Because of the equality $U \cap U_{\perp} = U \cap \text{Ker } g$ the above formula for the dimension of the subspace W can be written as follows:

$$\dim W = \dim U + \dim U_{\perp} - \dim(U \cap \text{Ker } g). \quad (3.1)$$

Let's compare (3.1) with the formula (2.6) from the theorem 2.5. This comparison yields $\dim W = \dim V$. Now, applying the item (3) of the theorem 4.5 from Chapter I, we get $W = V$. The theorem is proved. $\square$

THEOREM 3.2. *Let U be a subspace of a finite-dimensional space V regular with respect to a quadratic form g . If $U_{\perp} \neq \text{Ker } g$, then there exists a vector $\mathbf{v} \in U_{\perp}$ such that $g(\mathbf{v}) \neq 0$.*

PROOF. The proof is by contradiction. Assume that there is no vector $\mathbf{v} \in U_{\perp}$ such that $g(\mathbf{v}) \neq 0$. Then the numeric function $g(\mathbf{v})$ is identically zero in the subspace $U_{\perp}$. Due to the recovery formula (1.6) the numeric function $g(\mathbf{v}, \mathbf{w})$ is also identically zero for all $\mathbf{v}, \mathbf{w} \in U_{\perp}$.

Now let's apply the theorem 3.1 and expand an arbitrary vector $\mathbf{x} \in V$ into a sum of two vectors $\mathbf{x} = \mathbf{u} + \mathbf{w}$, where $\mathbf{u} \in U$ and $\mathbf{w} \in U_{\perp}$. Then for an arbitrary vector $\mathbf{v}$ of the subspace $U_{\perp}$ we derive

$$g(\mathbf{v}, \mathbf{x}) = g(\mathbf{v}, \mathbf{u} + \mathbf{w}) = g(\mathbf{v}, \mathbf{u}) + g(\mathbf{v}, \mathbf{w}) = 0 + 0 = 0.$$

The first summand $g(\mathbf{v}, \mathbf{u})$ in right hand side of the above equality is zero since the subspaces U and $U_{\perp}$ are orthogonal to each other. The second summand $g(\mathbf{v}, \mathbf{w})$ is zero due to our assumption in the beginning of the proof. Since $g(\mathbf{v}, \mathbf{x}) = 0$ for an arbitrary vector $\mathbf{x} \in V$, we get $\mathbf{v} \in \text{Ker } g$. But $\mathbf{v}$ is an arbitrary vector of the subspace $U_{\perp}$. Therefore, $U_{\perp} \subseteq \text{Ker } g$. The converse inclusion $\text{Ker } g \subseteq U_{\perp}$ is always valid. Hence, $U_{\perp} = \text{Ker } g$, which contradicts the hypothesis of the theorem. This contradiction means that the assumption, which we have made in the beginning of our proof, is not valid and, thus, it proves the existence of a vector $\mathbf{v} \in U_{\perp}$ such that $g(\mathbf{v}) \neq 0$. The theorem is proved. $\square$

THEOREM 3.3. *For any quadratic form g in a finite-dimensional vector space V there exists a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ such that the matrix of g is diagonal in this basis.*

PROOF. The case $g = 0$ is trivial. The matrix of the zero quadratic form g is purely zero in any basis. The square $n \times n$ matrix, which is purely zero, is obviously a diagonal matrix.

Suppose that $g \neq 0$. We shall prove the theorem by induction on the dimension of the space $\dim V = n$. In the case $n = 1$ the proposition of the theorem is trivial: any 1×1 matrix is a diagonal matrix.

Suppose that the theorem is valid for any quadratic form in any space of the dimension less than n . Let's consider the subspace $U = \text{Ker } g$. It is regular with respect to the form g and $U_{\perp} = V$. Therefore, we can apply the theorem 3.2. According to this theorem, there exists a vector $\mathbf{v}_0 \notin U$ such that $g(\mathbf{v}_0) \neq 0$. Let's consider the subspace W obtained by joining $\mathbf{v}_0$ to $U = \text{Ker } g$:

$$W = \text{Ker } g + \langle \mathbf{v}_0 \rangle = U \oplus \langle \mathbf{v}_0 \rangle. \quad (3.2)$$

This subspace W determines the following two cases: $W = V$ or $W \neq V$.

In the case $W = V$ we choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ in the kernel $\text{Ker } g$ and complete it by one additional vector $\mathbf{e}_{s+1} = \mathbf{v}_0$. As a result we get the basis in V . The matrix of the quadratic form g in this basis is a matrix almost completely filled with zeros, indeed, for $i = 1, \dots, s$ and $j = 1, \dots, s+1$ we have $g_{ij} = g_{ji} = g(\mathbf{e}_i, \mathbf{e}_j) = 0$ since $\mathbf{e}_i \in \text{Ker } g$. The only nonzero element is $g_{s+1, s+1}$, it is a diagonal element: $g_{s+1, s+1} = g(\mathbf{e}_{s+1}, \mathbf{e}_{s+1}) = g(\mathbf{v}_0) \neq 0$.

In the case $W \neq V$ we consider the intersection $W \cap W_\perp$. Let $\mathbf{w} \in W \cap W_\perp$. Then from (3.2) we derive $\mathbf{w} = \alpha \cdot \mathbf{v}_0 + \mathbf{u}$, where $\mathbf{u} \in \text{Ker } g$. Since $\mathbf{w}$ is a vector of W and simultaneously it is a vector of $W_\perp$, it should be orthogonal to itself with respect to the quadratic form g :

$$\begin{aligned} g(\mathbf{w}, \mathbf{w}) &= g(\alpha \cdot \mathbf{v}_0 + \mathbf{u}, \alpha \cdot \mathbf{v}_0 + \mathbf{u}) = \\ &= \alpha^2 g(\mathbf{v}_0, \mathbf{v}_0) + 2\alpha g(\mathbf{v}_0, \mathbf{u}) + g(\mathbf{u}, \mathbf{u}) = 0. \end{aligned} \quad (3.3)$$

But $\mathbf{u} \in \text{Ker } g$, therefore, $g(\mathbf{v}_0, \mathbf{u}) = 0$ and $g(\mathbf{u}, \mathbf{u}) = 0$, while $g(\mathbf{v}_0, \mathbf{v}_0) = g(\mathbf{v}_0) \neq 0$. Hence, from (3.3) we get $\alpha = 0$. This means that $\mathbf{w} = \mathbf{u} \in \text{Ker } g$. Thus, we have proved the inclusion $W \cap W_\perp \subseteq \text{Ker } g$, which means the regularity of the subspace W with respect to the quadratic form g .

Now let's apply the theorem 3.1. It yields the expansion $V = W + W_\perp$. Note that $\mathbf{v}_0 \in W$, but $\mathbf{v}_0 \notin W \cap W_\perp$. This follows from $g(\mathbf{v}_0, \mathbf{v}_0) \neq 0$. Hence, $\mathbf{v}_0 \notin W_\perp$ and $W_\perp \neq V$. This means that the dimension of the subspace $W_\perp$ is less than n . The formula (2.6) yield the exact value of this dimension

$$\dim W_\perp = \dim V + \dim(U \cap \text{Ker } g) - \dim U = n - 1.$$

Let's consider the restriction of g to the subspace $W_\perp$. We can apply the inductive hypothesis to g in $W_\perp$. Let $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$ be a basis of the subspace $W_\perp$ in which the matrix of the restriction of g to $W_\perp$ is diagonal:

$$g_{ij} = g_{ji} = g(\mathbf{e}_i, \mathbf{e}_j) = 0 \quad \text{for } i < j \leq n-1. \quad (3.4)$$

We complete this basis by one vector $\mathbf{e}_n = \mathbf{v}_0$. Since $\mathbf{v}_0 \notin W_\perp$ the extended system of vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ is linearly independent and, hence, is a basis of V . Let's find the matrix of the quadratic form g in the extended basis. For the elements in the extension of this matrix we obtain the relationships

$$g_{in} = g_{ni} = g(\mathbf{e}_i, \mathbf{e}_n) = 0 \quad \text{for } i < n. \quad (3.5)$$

They follow from the orthogonality of $\mathbf{e}_i$ and $\mathbf{e}_n$ in (3.5). Indeed, $\mathbf{e}_n \in W$ and $\mathbf{e}_i \in W_\perp$. The relationships (3.4) and (3.5) taken together mean that the matrix of the quadratic form g is diagonal in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. The inductive step is over and the theorem is completely proved. $\square$

Let g be a quadratic form in a finite dimensional space V and let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis in which the matrix of g is diagonal. Then the value of $g(\mathbf{v})$ can be calculated by formula (1.10). A part of the diagonal elements $g_{11}, \dots, g_{nn}$ can be equal to zero. Let's denote by s the number of such elements. We can renumerate the basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ so that

$$g_{11} = \dots = g_{ss} = 0. \quad (3.6)$$

The first s vectors of the basis, which correspond to the matrix elements (3.6), belong to the kernel of the form $\text{Ker } g$. Indeed, if $\mathbf{w} = \mathbf{e}_i$ for $i = 1, \dots, s$, then $g(\mathbf{v}, \mathbf{w}) = 0$ for all vectors $\mathbf{v} \in V$. This fact can be easily derived with the use of formulas (1.9).

Conversely, suppose that $\mathbf{w} \in \text{Ker } g$. Then for an arbitrary vector $\mathbf{v} \in V$ we have the following relationships:

$$g(\mathbf{v}, \mathbf{w}) = \sum_{i=1}^n \sum_{j=1}^n g_{ij} v^i w^j = \sum_{i=s+1}^n g_{ii} v^i w^i = 0.$$

Since $\mathbf{v} \in V$ is an arbitrary vector, the above equality should be fulfilled identically in $v^{s+1}, \dots, v^n$. But $g_{ii} \neq 0$ for $i \geq s+1$, therefore, $w^{s+1} = \dots = w^n = 0$. From these equalities for the vector $\mathbf{w}$ we derive

$$\mathbf{w} = w^1 \cdot \mathbf{e}_1 + \dots + w^s \cdot \mathbf{e}_s.$$

The conclusion is that any vector $\mathbf{w}$ of the kernel $\text{Ker } g$ can be expanded into a linear combination of the first s basis vectors. Hence, these basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_s$ form a basis in $\text{Ker } g$. The above considerations prove the following proposition that we present in the form of a theorem.

THEOREM 3.4. *The number of zeros on the diagonal of the matrix of a quadratic form g , brought to the diagonal form, is a geometric invariant of the form g . It does not depend on the method used for bringing this matrix to a diagonal form and coincides with the dimension of the kernel of the quadratic form: $s = \dim(\text{Ker } g)$.*

DEFINITION 3.2. The number $s = \dim(\text{Ker } g)$ is called the *zero inertia index* of a quadratic form g .

Let g be a quadratic form in a linear vector space over the field of complex numbers $\mathbb{C}$ such that its matrix is diagonal a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. Suppose that s is the zero inertia index of the quadratic form g . Without loss of generality we can assume that the first s basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_s$ form a basis in the kernel $\text{Ker } g$. We define the numbers $\gamma_1, \dots, \gamma_n$ by means of formula

$$\gamma_i = \begin{cases} 1 & \text{for } i \leq s, \\ \sqrt{g_{ii}} & \text{for } i > s. \end{cases} \quad (3.7)$$

Remember that for any complex number one can take its square root which is again a complex number. Complex numbers (3.7) are nonzero. We use them in order to construct the new basis:

$$\tilde{\mathbf{e}}_i = (\gamma_i)^{-1} \cdot \mathbf{e}_i, \quad i = 1, \dots, n. \quad (3.8)$$

The matrix of the quadratic form g in the new basis (3.8) is again a diagonal matrix. Indeed, we can explicitly calculate the matrix elements:

$$\tilde{g}_{ij} = g(\tilde{\mathbf{e}}_i, \tilde{\mathbf{e}}_j) = (\gamma_i \gamma_j)^{-1} g_{ij} = 0 \quad \text{for } i \neq j.$$

For the diagonal elements of the matrix of g we derive

$$\tilde{g}_{ii} = g(\tilde{\mathbf{e}}_i, \tilde{\mathbf{e}}_i) = (\gamma_i)^{-2} g_{ii} = \begin{cases} 0 & \text{for } i \leq s, \\ 1 & \text{for } i > s. \end{cases}$$

The matrix of the quadratic form g in the basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ has the following form which is used to be called the *canonic form* of the matrix of a quadratic form over the field of complex numbers $\mathbb{C}$:

$$\mathcal{G} = \left\| \begin{array}{ccccccc} 0 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & & & \\ & & & 1 & & & \\ & & & & \ddots & & \\ & & & & & 1 & \end{array} \right\| \begin{array}{l} \left. \vphantom{\begin{array}{c} 0 \\ \ddots \\ 0 \\ 1 \\ \ddots \\ 1 \end{array}} \right\} s \\ \left. \vphantom{\begin{array}{c} 1 \\ \ddots \\ 1 \end{array}} \right\} n-s \end{array} \quad (3.9)$$

The matrix $\mathcal{G}$ in (3.9) is a diagonal matrix, its diagonal is filled with s zeros and $n-s$ ones, where $s = \dim \text{Ker } g$.

In the case of a linear vector space over the field of real numbers $\mathbb{R}$ the canonic form of the matrix of a quadratic form is different from (3.9). Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis in which the matrix of g is diagonal. Diagonal elements of this matrix now is subdivided into three groups: zero elements, positive elements, and negative elements. If s is the number of zero elements and r is the number of positive elements, then remaining $n-s-r$ elements on the diagonal are negative numbers. Without loss of generality we can assume that the basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ are enumerated so that $g_{ii} = 0$ for $i = 1, \dots, s$ and $g_{ii} > 0$ for $i = s+1, \dots, s+r$. Then $g_{ii} < 0$ for $i = s+r+1, \dots, n$. In the field of reals we can take the square root only of non-negative numbers. Therefore, here we define $\gamma_1, \dots, \gamma_n$ a little bit differently than it was done in (3.7) for complex numbers:

$$\gamma_i = \begin{cases} 1 & \text{for } i \leq s, \\ \sqrt{|g_{ii}|} & \text{for } i > s. \end{cases} \quad (3.10)$$

By means of (3.10) we define new basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ using the formulas (3.9). Here is the matrix of the quadratic form g in this basis:

$$\mathcal{G} = \left\| \begin{array}{ccccccc} 0 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & & & \\ & & & 1 & & & \\ & & & & \ddots & & \\ & & & & & 1 & \\ & & & & & & -1 \\ & & & & & & \ddots \\ & & & & & & & -1 \end{array} \right\| \begin{array}{l} \left. \vphantom{\begin{array}{c} 0 \\ \ddots \\ 0 \\ 1 \\ \ddots \\ 1 \end{array}} \right\} s \\ \left. \vphantom{\begin{array}{c} 1 \\ \ddots \\ 1 \end{array}} \right\} r_p \\ \left. \vphantom{\begin{array}{c} -1 \\ \ddots \\ -1 \end{array}} \right\} r_n \end{array} \quad (3.11)$$

DEFINITION 3.3. The formula (3.11) defines the *canonic form* of the matrix of a quadratic form g in a space over the real numbers $\mathbb{R}$. The integers r_p and r_n that determine the number of plus ones and the number of minus ones on the diagonal of the matrix (3.10) are called the *positive inertia index* and the *negative inertia index* of the quadratic form g respectively.

THEOREM 3.5. The positive and the negative inertia indices r_p and r_n of a quadratic form g in a space over the field of real numbers $\mathbb{R}$ are geometric invariants of g . They do not depend on a particular way how the matrix of g was brought to the diagonal form.

PROOF. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis of a space V in which the matrix of g has the canonic form (3.11). Let's consider the following subspaces:

$$U_+ = \langle \mathbf{e}_1, \dots, \mathbf{e}_{s+r_p} \rangle, \quad U_- = \langle \mathbf{e}_{s+r_p+1}, \dots, \mathbf{e}_n \rangle. \quad (3.12)$$

The intersection of U_+ and U_- is trivial, $\dim U_+ = s + r_p$, $\dim U_- = r_n$, and for their sum we have $U_+ \oplus U_- = V$.

Let's take a vector $\mathbf{v} \in U_+$. The value of the quadratic form g for that vector is determined by the matrix (3.11) according to the formula (1.10):

$$g(\mathbf{v}) = \sum_{i=s+1}^{s+r_p} (v^i)^2.$$

The sum of squares in the right hand side of this equality is a non-negative quantity, i. e. $g(\mathbf{v}) \geq 0$ for all $\mathbf{v} \in U_+$.

Now let's take a vector $\mathbf{v} \in U_-$. For this vector the formula (1.10) is written as

$$g(\mathbf{v}) = \sum_{i=s+r_p+1}^n -(v^i)^2.$$

If $\mathbf{v} \neq \mathbf{0}$, then at least one summand in right hand side is nonzero. Hence, $g(\mathbf{v}) < 0$ for all nonzero vectors of the subspace U_- .

Suppose that $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ is some other basis in which the matrix of g has the canonic form. Denote by $\tilde{s}$, $\tilde{r}_p$, and $\tilde{r}_n$ the inertia indices of g in this basis. The zero inertia indices in both bases are the same $s = \tilde{s}$ since they are determined by the kernel of g : $s = \dim(\text{Ker } g)$ and $\tilde{s} = \dim(\text{Ker } g)$.

Let's prove the coincidence of the positive and the negative inertia indices in two bases. For this purpose we consider the subspaces $\tilde{U}_+$ and $\tilde{U}_-$ determined by the relationships of the form (3.12) but for the «wavy» basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$. If we assume that $r_p \neq \tilde{r}_p$, then $r_p > \tilde{r}_p$ or $r_p < \tilde{r}_p$. For the sake of certainty suppose that $r_p > \tilde{r}_p$. Then we calculate the dimensions of U_+ and $\tilde{U}_-$:

$$\dim U_+ = s + r_p, \quad \dim \tilde{U}_- = \tilde{r}_n = n - s - \tilde{r}_p.$$

For the sum of dimensions of these two subspaces U_+ and $\tilde{U}_-$ we get the equality $\dim U_+ + \dim \tilde{U}_- = n + (r_p - \tilde{r}_p)$. Due to the above assumption $r_p > \tilde{r}_p$ we derive

$$\dim U_+ + \dim \tilde{U}_- > \dim V \quad (3.13)$$

From the natural inclusion $U_+ + \tilde{U}_- \subseteq V$ we get $\dim(U_+ + \tilde{U}_-) \leq \dim V$. Using this estimate together with the inequality (3.13) and applying the theorem 6.4 of Chapter I to them, we derive $\dim(U_+ \cap \tilde{U}_-) > 0$. Hence, the intersection $U_+ \cap \tilde{U}_-$ is nonzero, it contains a nonzero vector $\mathbf{v} \in U_+ \cap \tilde{U}_-$. From the conditions $\mathbf{v} \in U_+$ and $\mathbf{v} \in U_-$ we obtain two inequalities

$$g(\mathbf{v}) \geq 0, \quad g(\mathbf{v}) < 0$$

contradicting each other. This contradiction shows that our assumption $r_p \neq \tilde{r}_p$ is not valid and the inertia indices r_p and $\tilde{r}_p$ do coincide. From $r_p = \tilde{r}_p$ and $s = \tilde{s}$ then we derive $r_n = \tilde{r}_n$. The theorem is proved. $\square$

DEFINITION 3.4. The total set of inertia indices is called the *signature* of a quadratic form. In the case of a quadratic form in complex space ($\mathbb{K} = \mathbb{C}$) the signature is formed by two numbers $(s, n - s)$, in the case of real space ($\mathbb{K} = \mathbb{R}$) it is formed by three numbers (s, r_p, r_n) .

In the case of a linear space over the field of rational numbers $\mathbb{K} = \mathbb{Q}$ we can also diagonalize the matrix of a quadratic form and subdivide the diagonal elements into three parts: positive, negative, and zero elements. This determines the numbers s , r_p , and r_n , which are geometric invariants of g , and we can define its signature.

However, in the case $\mathbb{K} = \mathbb{Q}$ we cannot reduce the nonzero diagonal elements to plus ones and minus ones only. Therefore, the number of geometric invariants in this case is greater than 3. We shall not look for the complete set of geometric invariants of a quadratic form in the case $\mathbb{K} = \mathbb{Q}$ and we shall not construct their theory since this would lead us to the number theory toward the problems of divisibility, primality, factorization of integers, etc.

§ 4. Positive quadratic forms. Sylvester's criterion.

In this section we consider quadratic forms in linear vector spaces over the field of real numbers $\mathbb{R}$. However, almost all results of this section remain valid for quadratic forms in rational vector spaces as well.

DEFINITION 4.1. A quadratic form g in a space V over the field $\mathbb{R}$ is called a *positive* form if $g(\mathbf{v}) > 0$ for any nonzero vector $\mathbf{v} \in V$.

THEOREM 4.1. A quadratic form g in a finite-dimensional space V is positive if and only if the numbers s and r_n in its signature (s, r_p, r_n) are equal to zero.

PROOF. Let g be a positive quadratic form and let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis in which the matrix of g has the canonic form (3.11). If $s \neq 0$ then for the basis vector $\mathbf{e}_1 \neq 0$ we would get $g(\mathbf{e}_1) = g_{11} = 0$, which would contradict the positivity of g . If $r_n \neq 0$, then for the basis vector $\mathbf{e}_n \neq 0$ we would get $g(\mathbf{e}_n) = g_{nn} = -1$, which would also contradict the positivity of g . Hence, $s = r_n = 0$.

Now, conversely, let $s = r_n = 0$. Then in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$, in which the matrix of g has the form (3.11), its value $g(\mathbf{v})$ is the sum of squares

$$g(v) = (v^1)^2 + \dots + (v^n)^2,$$

where $v^1, \dots, v^n$ are coordinates of a vector $\mathbf{v}$. This formula follows from the formula (1.10). For a nonzero vector at least one of its coordinates is nonzero. Hence, $g(\mathbf{v}) > 0$. This proves the positivity of g and thus completes the proof of the theorem in whole. $\square$

The condition $s = \text{Ker } g$ obtained in the theorem 3.4 and the condition $s = 0$ mean that a positive form g in a finite-dimensional space V is non-degenerate: $\text{Ker } g = \{\mathbf{0}\}$. This fact is valid for a form in an infinite-dimensional space as well.

THEOREM 4.2. *Any positive quadratic form g is non-degenerate.*

PROOF. If $\text{Ker } g \neq \{\mathbf{0}\}$ then there is a nonzero vector $\mathbf{v} \in \text{Ker } g$. The vector $\mathbf{v}$ of the kernel is orthogonal to all vectors of the space V . Hence, it is orthogonal to itself: $g(\mathbf{v}) = g(\mathbf{v}, \mathbf{v}) = 0$. If so, this fact contradicts the positivity of the form g . Therefore, any positive form g should be non-degenerate. $\square$

THEOREM 4.3. *Any subspace $U \subset V$ is regular with respect to a positive quadratic form g in a linear vector space V .*

PROOF. Since the kernel $\text{Ker } g$ of a positive form g is zero, the regularity of a subspace U with respect to g is equivalent to the equality $U \cap U_\perp = \{\mathbf{0}\}$ (see definition 3.1). Let's prove this equality. Let $\mathbf{v}$ be an arbitrary vector of the intersection $U \cap U_\perp$. From $\mathbf{v} \in U_\perp$ we derive that it is orthogonal to all vectors of U . Hence, it is also orthogonal to itself since $\mathbf{v} \in U$. Therefore, $g(\mathbf{v}) = g(\mathbf{v}, \mathbf{v}) = 0$. Due to positivity of g the equality $g(\mathbf{v}) = 0$ holds only for the zero vector $\mathbf{v} = \mathbf{0}$. Thus, we get $U \cap U_\perp = \{\mathbf{0}\}$. The theorem is proved. $\square$

THEOREM 4.4. *For any subspace $U \subset V$ and for any positive quadratic form g in a finite-dimensional space V there is an expansion $V = U \oplus U_\perp$.*

PROOF. The expansion $V = U + U_\perp$ follows from the theorem 3.1. We need only to prove that the sum in this expansion is a direct sum. For the sum of the dimensions of U and $U_\perp$ from the theorem 2.5 due to the triviality of the kernel $\text{Ker } g = \{\mathbf{0}\}$ of a positive quadratic form g we derive

$$\dim U + \dim U_\perp = \dim V.$$

Due to this equality in order to complete the proof it is sufficient to apply the theorem 6.3 of Chapter I. $\square$

Let g be a quadratic form in a finite-dimensional space V over the field of real numbers $\mathbb{R}$. Let's choose an arbitrary basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in V and then let's construct the matrix of the quadratic form g :

$$\mathcal{G} = \begin{vmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{n1} & \cdots & g_{nn} \end{vmatrix} \quad (4.1)$$

Let's delete the last $n - k$ columns and the last $n - k$ rows in the above matrix (4.1). The determinant of the matrix thus obtained is called the k -th principal

minor of the matrix $\mathcal{G}$. We denote this determinant by M_k :

$$M_k = \det \begin{vmatrix} g_{11} & \cdots & g_{1k} \\ \vdots & \ddots & \vdots \\ g_{k1} & \cdots & g_{kk} \end{vmatrix} \quad (4.2)$$

The n -th principal minor M_n coincides with the determinant of the matrix $\mathcal{G}$.

THEOREM 4.5. *Let g be a positive quadratic form in a finite-dimensional space V . Then the determinant of the matrix of g in an arbitrary basis of V is positive.*

PROOF. For the beginning we consider a canonic basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in which the matrix of g has the canonic form (3.11). According to the theorem 4.1, the matrix of a positive quadratic form g in a canonic basis is the unit matrix. Hence, its determinant is equal to unity and thus it is positive: $\det \mathcal{G} = 1 > 0$.

Now let $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ be an arbitrary basis and let S be the transition matrix for passing from $\mathbf{e}_1, \dots, \mathbf{e}_n$ to $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$. Applying the formula (1.12), we get

$$\det \tilde{\mathcal{G}} = \det S^{\text{tr}} (\det \mathcal{G}) \det S = (\det S)^2.$$

In a linear vector space V over the real numbers $\mathbb{R}$ the elements of any transition matrix S are real numbers. Its determinant is also a nonzero real number. Therefore, $(\det S)^2$ is a positive number. The theorem is proved. $\square$

Now again let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be an arbitrary basis of V and let g_{ij} be the matrix of a positive quadratic form g in this basis. Let's consider the subspace

$$U_k = \langle \mathbf{e}_1, \dots, \mathbf{e}_k \rangle.$$

Let's denote by h_k the restriction of g to the subspace U_k . The matrix of the form h_k in the basis $\mathbf{e}_1, \dots, \mathbf{e}_k$ coincides with upper left diagonal block in the matrix of the initial form g . This is the very block that determines the k -th principal minor M_k in the formula (4.2). It is clear that the restriction of a positive form g to any subspace is again a positive quadratic form. Therefore, we can apply the theorem 4.5 to the form h_k . This yields $M_k > 0$.

Conclusion: the positivity of all principal minors (4.2) is a necessary condition for the positivity of a quadratic form g itself. As appears, this condition is a sufficient condition as well. This fact is known as *Sylvester's criterion*.

THEOREM 4.6 (SYLVESTER). *A quadratic form g in a finite-dimensional space V is positive if and only if all principal minors of its matrix are positive.*

PROOF. The positivity of g implies the positivity of all principal minors in its matrix. This fact is already proved. Let's prove the converse proposition. Suppose that all diagonal minors (4.2) in the matrix of a quadratic form g are positive. We should prove that g is positive. The proof is by induction on $n = \dim V$.

The basis of the induction in the case $\dim V = 1$ is obvious. Here the matrix of g consists of the only element g_{11} that coincides with the only principal minor: $g_{11} = M_1$. The value $g(\mathbf{v})$ in one-dimensional space is determined by the only coordinate of a vector $\mathbf{v}$ according to the formula $g(\mathbf{v}) = g_{11} (v^1)^2$. Therefore $M_1 > 0$ implies the positivity of the form g .

Suppose that the proposition we are going to prove is valid for a quadratic form in any space of the dimension less than $n = \dim V$. Let g_{ij} be the matrix of our quadratic form g in some basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of V . Let's denote

$$U = \langle \mathbf{e}_1, \dots, \mathbf{e}_{n-1} \rangle.$$

Denote by h the restriction of the form g to the subspace U of the dimension $n-1$. The matrix elements h_{ij} in the matrix of h calculated in the basis $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$ coincide with corresponding elements in the matrix of the initial form: $h_{ij} = g_{ij}$. Therefore, the minors $M_1, \dots, M_{n-1}$ can be calculated by means of the matrix h_{ij} . Due to the positivity of these minors, applying the inductive hypothesis, we find that h is a positive quadratic form in U .

Let $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_{n-1}$ be a basis in which the matrix of the form h has the canonic form (3.11). Applying the theorem 4.1 to the form h , we conclude that the matrix $\tilde{h}_{ij}$ in the canonic basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_{n-1}$ is the unit matrix. Let's complete the basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_{n-1}$ of the subspace U by the vector $\mathbf{e}_n \notin U$. As a result we get the basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_{n-1}, \mathbf{e}_n$ in which the matrix of g has the form

$$\mathcal{G}_1 = \begin{vmatrix} 1 & \dots & 0 & \tilde{g}_{1n} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & \tilde{g}_{n-1n} \\ \tilde{g}_{n1} & \dots & \tilde{g}_{nn-1} & g_{nn} \end{vmatrix}. \quad (4.3)$$

The passage from the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ to the basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_{n-1}, \mathbf{e}_n$ is described by a blockwise-diagonal matrix S of the form

$$S_1 = \begin{vmatrix} S_1^1 & \dots & S_n^1 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ S_1^{n-1} & \dots & S_n^{n-1} & 0 \\ 0 & \dots & 0 & 1 \end{vmatrix}. \quad (4.4)$$

The formula (1.12) relates the matrix (4.3) with the matrix $\mathcal{G}$ of the quadratic form g in the initial basis: $\mathcal{G}_1 = S^{\text{tr}} \mathcal{G} S$. From this formula we derive

$$\det \mathcal{G}_1 = \det \mathcal{G} (\det S)^2 = M_n (\det S)^2. \quad (4.5)$$

Due to the above formula (4.5) the positivity of the principal minor $M_n = \det \mathcal{G}$ in the initial matrix (4.1) implies the positivity of the determinant of the matrix (4.3), i.e. $\det \mathcal{G}_1 > 0$.

Let's calculate the determinant of the matrix (4.3) explicitly. For this purpose we multiply the first column of this matrix by $\tilde{g}_{1n}$ and subtract it from the last column. Then we multiply the second column by $\tilde{g}_{2n}$ and subtract it from the last one. We produce such an operation repeatedly for each of the first $n-1$ columns of the matrix (4.3). From the course of algebra we know that such transformations do not change the determinant of a matrix. In present case they simplify the matrix (4.3) bringing it to a lower-triangular form. Therefore, we can calculate

the determinant of the matrix (4.3) in explicit form:

$$\det \mathcal{G}_1 = \det \begin{vmatrix} 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \\ \tilde{g}_{n1} & \dots & \tilde{g}_{nn-1} & \tilde{g}_{nn} \end{vmatrix} = \tilde{g}_{nn}. \quad (4.6)$$

The element $\tilde{g}_{nn}$ in the transformed matrix is given by the formula

$$\tilde{g}_{nn} = g_{nn} - \sum_{i=1}^{n-1} g_{ni} g_{in} = g_{nn} - \sum_{i=1}^{n-1} (g_{in})^2. \quad (4.7)$$

The matrix of the quadratic form g in the basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_{n-1}, \mathbf{e}_n$ is close to the diagonal matrix. Let's complete the process of diagonalization replacing the vector $\mathbf{e}_n$ by the vector $\tilde{\mathbf{e}}_n \notin U$ such that

$$\tilde{\mathbf{e}}_n = \mathbf{e}_n - \sum_{i=1}^{n-1} g_{in} \cdot \tilde{\mathbf{e}}_i.$$

The passage from $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_{n-1}, \mathbf{e}_n$ to $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ changes only the last basis vector. Therefore, the unit diagonal block in the matrix (4.3) remains unchanged. For non-diagonal elements $g(\tilde{\mathbf{e}}_k, \tilde{\mathbf{e}}_n)$ in the new basis we have

$$g(\tilde{\mathbf{e}}_k, \tilde{\mathbf{e}}_n) = \tilde{g}_{kn} - \sum_{i=1}^{n-1} g_{in} g(\tilde{\mathbf{e}}_k, \tilde{\mathbf{e}}_i) = \tilde{g}_{kn} - \sum_{i=1}^{n-1} g_{in} \tilde{h}_{ki} = 0.$$

The equality $g(\tilde{\mathbf{e}}_k, \tilde{\mathbf{e}}_n) = 0$ in the above formula is due to the fact that the matrix of the restricted form h in its canonic basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_{n-1}$ is the unit matrix. For the diagonal element $g(\tilde{\mathbf{e}}_n, \tilde{\mathbf{e}}_n)$ from this fact we derive

$$g(\tilde{\mathbf{e}}_n, \tilde{\mathbf{e}}_n) = g_{nn} - \sum_{i=1}^{n-1} \sum_{k=1}^{n-1} g_{in} g_{kn} h_{ik} = g_{nn} - \sum_{i=1}^{n-1} (g_{in})^2.$$

Comparing this expression with (4.7), we find that $g(\tilde{\mathbf{e}}_n, \tilde{\mathbf{e}}_n) = \tilde{g}_{nn}$. Thus, the matrix of g in the basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ is a diagonal matrix of the form

$$\mathcal{G}_2 = \begin{vmatrix} 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \\ 0 & \dots & 0 & \tilde{g}_{nn} \end{vmatrix}. \quad (4.8)$$

Combining (4.5) and (4.6), for the element $\tilde{g}_{nn}$ in (4.8) we get $\tilde{g}_{nn} = M_n (\det S)^2$. Since the principal minor M_n of the initial matrix (4.1) is positive, we find that $\tilde{g}_{nn}$ in (4.8) is also positive. Hence, g is a positive quadratic form. Thus, we have completed the inductive step and have proved the theorem in whole. $\square$

EUCLIDEAN SPACES.

§ 1. The norm and the scalar product. The angle between vectors. Orthonormal bases.

DEFINITION 1.1. A *Euclidean vector space* is a linear vector space V over the field of reals $\mathbb{R}$ which is equipped with some fixed positive quadratic form g .

Let (V, g) be a Euclidean vector space. There many positive quadratic forms in the linear vector space V , however, only one of them is associated with V so that it defines the structure of Euclidean space in V . Two Euclidean vector spaces (V, g_1) and (V, g_2) with $g_1 \neq g_2$ coincide as linear vector spaces, but they are different when considered as Euclidean vector spaces.

The structure of the Euclidean vector space (V, g) is associated with a special terminology and special notations. The value of the quadratic form $g(\mathbf{v})$ is non-negative. The square root of $g(\mathbf{v})$ is called the *norm* or the *length* of a vector $\mathbf{v}$. The norm of a vector $\mathbf{v}$ is denoted as follows:

$$|\mathbf{v}| = \sqrt{g(\mathbf{v})}. \quad (1.1)$$

The quadratic form $g(\mathbf{v})$ produces the bilinear form $g(\mathbf{v}, \mathbf{w})$ determined by the recovery formula (1.6) of Chapter IV. The value of that bilinear form is called the *scalar product* of two vectors $\mathbf{v}$ and $\mathbf{w}$. The scalar product is denoted as follows:

$$(\mathbf{v} | \mathbf{w}) = g(\mathbf{v}, \mathbf{w}). \quad (1.2)$$

Due to the notation (1.1) and (1.2), when dealing with some fixed Euclidean space (V, g) , we can omit the symbol g at all.

The scalar product (1.2) is defined for a pair of two vectors $\mathbf{v}, \mathbf{w} \in V$. It is quite different from the scalar product (1.8) of Chapter III, which is defined for a pair of a vector and a covector. The scalar product (1.2) of a Euclidean vector space possesses the following properties:

- (1) $(\mathbf{v}_1 + \mathbf{v}_2 | \mathbf{w}) = (\mathbf{v}_1 | \mathbf{w}) + (\mathbf{v}_2 | \mathbf{w})$ for all $\mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in V$;
- (2) $(\alpha \cdot \mathbf{v} | \mathbf{w}) = \alpha (\mathbf{v} | \mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in V$ and for all $\alpha \in \mathbb{R}$;
- (3) $(\mathbf{v} | \mathbf{w}_1 + \mathbf{w}_2) = (\mathbf{v} | \mathbf{w}_1) + (\mathbf{v} | \mathbf{w}_2)$ for all $\mathbf{w}_1, \mathbf{w}_2, \mathbf{v} \in V$;
- (4) $(\mathbf{v} | \alpha \cdot \mathbf{w}) = \alpha (\mathbf{v} | \mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in V$ and for all $\alpha \in \mathbb{R}$;
- (5) $(\mathbf{v} | \mathbf{w}) = (\mathbf{w} | \mathbf{v})$ for all $\mathbf{v}, \mathbf{w} \in V$;
- (6) $|\mathbf{v}|^2 = (\mathbf{v} | \mathbf{v}) \geq 0$ for all $\mathbf{v} \in V$ and $|\mathbf{v}| = 0$ implies $\mathbf{v} = \mathbf{0}$.

The properties (1)-(4) reflect the bilinearity of the form g in (1.2). They are analogous to that of the scalar product of a vector and a covector (see formulas (1.9) in Chapter III).

The properties (5) and (6) have no such analogs. But they are the very properties that make the scalar product (1.2) a generalization of the scalar product of 3-dimensional geometric vectors.

THEOREM 1.1. *The following two additional properties of the scalar product (1.2) are derived from the properties (1)-(6):*

- (7) $|(\mathbf{v}, \mathbf{w})| \leq |\mathbf{v}| |\mathbf{w}|$ for all $\mathbf{v}, \mathbf{w} \in V$;
- (8) $|\mathbf{v} + \mathbf{w}| \leq |\mathbf{v}| + |\mathbf{w}|$ for all $\mathbf{v}, \mathbf{w} \in V$.

The property (7) is known as the *Cauchy-Bunyakovsky-Schwarz inequality*, while the property (8) is called the *triangle inequality*.

PROOF. In order to prove the inequality (7) we choose two arbitrary nonzero vectors $\mathbf{v}, \mathbf{w} \in V$ and consider the numeric function $f(\alpha)$ of a numeric argument α defined by the following explicit formula:

$$f(\alpha) = |\mathbf{v} + \alpha \cdot \mathbf{w}|^2. \quad (1.3)$$

Using the properties (1)-(6) we find that $f(\alpha)$ is a polynomial of degree two:

$$\begin{aligned} f(\alpha) &= |\mathbf{v} + \alpha \cdot \mathbf{w}|^2 = (\mathbf{v} + \alpha \cdot \mathbf{w} | \mathbf{v} + \alpha \cdot \mathbf{w}) = \\ &= (\mathbf{v} | \mathbf{v}) + 2\alpha (\mathbf{v} | \mathbf{w}) + \alpha^2 (\mathbf{w} | \mathbf{w}). \end{aligned}$$

The function (1.3) has a lower bound: $f(\alpha) \geq 0$. This follows from the property (6). Let's calculate the minimum point of the function $f(\alpha)$ by equating its derivative $f'(\alpha)$ to zero. This yields the following equation:

$$f'(\alpha) = 2(\mathbf{v} | \mathbf{w}) + 2\alpha (\mathbf{w} | \mathbf{w}) = 0.$$

Solving this equation, we find $\alpha_{\min} = -(\mathbf{v} | \mathbf{w})/(\mathbf{w} | \mathbf{w})$. Now let's write the condition $f(\alpha) \geq 0$ for the minimal value of the function $f(\alpha)$:

$$f_{\min} = f(\alpha_{\min}) = \frac{|\mathbf{v}|^2 |\mathbf{w}|^2 - (\mathbf{v} | \mathbf{w})^2}{|\mathbf{w}|^2} \geq 0. \quad (1.4)$$

The denominator of the fraction (1.4) is positive, therefore, from the inequality (1.4) we easily derive the property (7).

In order to prove the property (8) we consider the square of the norm for the vector $\mathbf{v} + \mathbf{w}$. For this quantity we derive

$$|\mathbf{v} + \mathbf{w}|^2 = (\mathbf{v} + \mathbf{w} | \mathbf{v} + \mathbf{w}) = |\mathbf{v}|^2 + 2(\mathbf{v} | \mathbf{w}) + |\mathbf{w}|^2. \quad (1.5)$$

Applying the property (8), which is already proved, for the right hand side of the equality (1.5) we get the following estimate:

$$|\mathbf{v}|^2 + 2(\mathbf{v} | \mathbf{w}) + |\mathbf{w}|^2 \leq |\mathbf{v}|^2 + 2|\mathbf{v}| |\mathbf{w}| + |\mathbf{w}|^2 = (|\mathbf{v}| + |\mathbf{w}|)^2.$$

From the relationship (1.5) and from the above inequality we derive the other inequality $|\mathbf{v} + \mathbf{w}|^2 \leq (|\mathbf{v}| + |\mathbf{w}|)^2$. Now the property (8) is derived by taking the square root of both sides of this inequality. This operation is correct since $y = \sqrt{x}$ is an increasing function of the real semiaxis $[0, +\infty)$. The theorem is proved. $\square$

Due to the analogy of (1.2) and the scalar product of geometric vectors and due to the Cauchy-Bunyakovsky-Schwarz inequality $|(\mathbf{v}, \mathbf{w})| \leq |\mathbf{v}| |\mathbf{w}|$ we can introduce the concept of an *angle between vectors* in a Euclidean vector space.

DEFINITION 1.2. The number φ from the interval $0 \leq \varphi \leq \pi$, which is determined by the following implicit formula

$$\cos(\varphi) = \frac{(\mathbf{v} | \mathbf{w})}{|\mathbf{v}| |\mathbf{w}|}, \quad (1.6)$$

is called the *angle* between two nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ in a Euclidean space V .

Due to the property (7) from the theorem 1.1 the modulus of the fraction in left hand side of (1.6) is not greater than 1. Therefore, the formula (1.6) is correct. It determines the unique number φ from the specified interval $0 \leq \varphi \leq \pi$.

DEFINITION 1.3. Two vectors $\mathbf{v}$ and $\mathbf{w}$ in a Euclidean space V are called *orthogonal vectors* if they form a right angle ($\varphi = \pi/2$).

The definition 1.3 applies only to nonzero vectors $\mathbf{v}$ and $\mathbf{w}$. The definition 2.1 of Chapter IV is more general. Let's reformulate for the case of Euclidean spaces.

DEFINITION 1.4. Two vectors $\mathbf{v}$ and $\mathbf{w}$ in a Euclidean space V are called *orthogonal vectors* if their scalar product is zero: $(\mathbf{v} | \mathbf{w}) = 0$.

For nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ these two definitions 1.3 and 1.4 are equivalent.

Let $\mathbf{v}_1, \dots, \mathbf{v}_m$ be a system of vectors in a Euclidean space (V, g) . The matrix g_{ij} composed by the mutual scalar products of these vectors

$$g_{ij} = (\mathbf{v}_i | \mathbf{v}_j), \quad (1.7)$$

is called the *Gram matrix* of the system of vectors $\mathbf{v}_1, \dots, \mathbf{v}_m$.

THEOREM 1.2. A system of vectors $\mathbf{v}_1, \dots, \mathbf{v}_m$ in a Euclidean space is linearly dependent if and only if the determinant of their Gram matrix is equal to zero.

PROOF. Suppose that the vectors $\mathbf{v}_1, \dots, \mathbf{v}_m$ are linearly dependent. Then there is a nontrivial linear combination of these vectors which is equal to zero:

$$\alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_m \cdot \mathbf{v}_m = \mathbf{0}. \quad (1.8)$$

Using the coefficients of the linear combination (1.8), we construct the following expression with the components of Gram matrix (1.7):

$$\sum_{j=1}^m g_{ij} \alpha_j = \sum_{j=1}^m (\mathbf{v}_i | \mathbf{v}_j) \alpha_j = (\mathbf{v}_i | \alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_m \cdot \mathbf{v}_m) = (\mathbf{v}_i | \mathbf{0}) = 0.$$

Since i is a free index running over the interval of integer numbers from 1 to m , this formula means that the columns of Gram matrix g_{ij} are linearly dependent. Hence, its determinant is equal to zero (this fact is known from the course of general algebra).

Conversely, assume that the determinant of the Gram matrix (1.7) is equal to zero. Then the columns of this matrix are linearly dependent and, hence, there is a nontrivial linear combination of them that is equal to zero:

$$\sum_{j=1}^m g_{ij} \alpha_j = 0. \quad (1.9)$$

Let's denote $\mathbf{v} = \alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_m \cdot \mathbf{v}_m$. Then consider the following double sum, which is obviously equal to zero due to the equality (1.9):

$$\begin{aligned} 0 &= \sum_{i=1}^m \sum_{j=1}^m \alpha_i g_{ij} \alpha_j = \sum_{i=1}^m \alpha_i (\mathbf{v}_i | \alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_m \cdot \mathbf{v}_m) = \\ &= (\alpha_1 \cdot \mathbf{v}_1 + \dots + \alpha_m \cdot \mathbf{v}_m | \mathbf{v}) = (\mathbf{v} | \mathbf{v}) = |\mathbf{v}|^2. \end{aligned}$$

Thus, we get $|\mathbf{v}|^2 = 0$ and, using the positivity of the basic quadratic form g of the Euclidean space V , we derive $\mathbf{v} = \mathbf{0}$. Since $\mathbf{v} = \mathbf{0}$, we get the nontrivial linear combination of the form (1.8), which is equal to zero. Hence, our vector $\mathbf{v}_1, \dots, \mathbf{v}_m$ are linearly dependent. $\square$

Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis in a finite-dimensional Euclidean vector space (V, g) . Let's consider the Gram matrix of this basis. Knowing the components of the Gram matrix, we can calculate the norm of vectors (1.1) and the scalar product of vectors (1.2) through their coordinates:

$$|\mathbf{v}|^2 = \sum_{i=1}^n \sum_{j=1}^n g_{ij} v^i v^j, \quad (\mathbf{v} | \mathbf{w}) = \sum_{i=1}^n \sum_{j=1}^n g_{ij} v^i w^j. \quad (1.10)$$

A basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in a Euclidean space V is called an *orthonormal basis* if the Gram matrix for the basis vectors is the unit matrix:

$$g_{ij} = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases} \quad (1.11)$$

If the condition (1.11) is not fulfilled, then the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ is called a *skew-angular basis*. In an orthonormal basis the vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ are unit vectors orthogonal to each other. This simplifies the formulas (1.10) substantially:

$$|\mathbf{v}|^2 = \sum_{i=1}^n (v^i)^2, \quad (\mathbf{v} | \mathbf{w}) = \sum_{i=1}^n v^i w^i. \quad (1.12)$$

Orthonormal bases do exist. Due to (1.2) and (1.7) we know that the Gram matrix of the basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ is the matrix of the quadratic g in this basis. The theorem 3.3 of Chapter IV says that there exists a basis in which the matrix of g has its canonic form (see (3.11) in Chapter IV). Since g is a positive quadratic form, its matrix in a canonic form is the unit matrix (see theorem 4.1 in Chapter IV).

The theorem 4.8 on completing the basis of a subspace formulated in Chapter I has its analog for orthonormal bases.

THEOREM 1.3. *Let $\mathbf{e}_1, \dots, \mathbf{e}_s$ be an orthonormal basis in a subspace U of a finite-dimensional Euclidean space (V, g) . Then it can be completed up to an orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in V .*

PROOF. Let's consider the orthogonal complement $U_\perp$ of the subspace U . According to the theorem 4.4 of Chapter IV, the subspaces U and $U_\perp$ define the expansion of the space V into a direct sum:

$$V = U \oplus U_\perp.$$

The subspace $U_\perp$ inherits the structure of a Euclidean space from V . Let's choose an orthonormal basis $\mathbf{e}_{s+1}, \dots, \mathbf{e}_n$ in $U_\perp$ and then join together two bases of U and $U_\perp$. As a result we get the basis in V (see theorem 6.3 of Chapter I). The vectors of this basis are unit vectors by their length and they are orthogonal to each other. Hence, this is an orthonormal basis completing the initial basis $\mathbf{e}_1, \dots, \mathbf{e}_s$ of the subspace U . $\square$

Let $\mathbf{e}_1, \dots, \mathbf{e}_s$ and $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_s$ be two orthonormal bases and let S be the transition. The Gram matrices of these two bases are the unit matrices. Therefore, applying the formulas (1.12) of Chapter IV, for the transition matrix S we derive

$$S^{\text{tr}} S = 1, \quad S^{-1} = S^{\text{tr}}. \quad (1.13)$$

Note that a square matrix S satisfying the above relationships (1.13) is called an *orthogonal* matrix.

From the relationships (1.13) for the determinant of an orthogonal matrix we get: $(\det S)^2 = 1$. Therefore, orthogonal matrices are subdivided into two types: matrices with positive determinant $\det S = 1$ and those with negative determinant $\det S = -1$. This subdivision is related to the concept of *orientation*. All bases in a linear vector space over the field of real numbers $\mathbb{R}$ (not necessarily a Euclidean space) can be subdivided into two sets which can be called «left bases» and «right bases». The transition matrix for passing from a left basis to a left basis or for passing from a right basis to another right basis is a matrix with positive determinant — it does not change the orientation. The transition matrix for passing from a left basis to a right basis or, conversely, from a right basis to a left basis is a matrix with negative determinant. Such a transition matrix changes the orientation of a basis. We say that a linear vector space V over the field of real numbers $\mathbb{R}$ is equipped with the *orientation* if there is some mechanism to distinguish one of two types of bases in V .

§ 2. Quadratic forms in a Euclidean space. Diagonalization of a pair of quadratic forms.

Let (V, g) be a Euclidean vector space and let φ be a quadratic form in V . For such a form φ we define the following ratio:

$$\mu(\mathbf{v}) = \frac{|\varphi(\mathbf{v})|}{|\mathbf{v}|^2}. \quad (2.1)$$

The number $\mu(v)$ in (2.1) is a real non-negative number. Note that $\mu(\alpha \cdot \mathbf{v}) = \mu(\mathbf{v})$ for any nonzero $\alpha \in \mathbb{R}$. Therefore, we can assume $\mathbf{v}$ in (2.1) to be a unit vector.

Let's denote by $\|\varphi\|$ the least upper bound of $\mu(v)$ for all unit vectors (such vectors sweep out the *unit sphere* in the Euclidean space V):

$$\|\varphi\| = \sup_{|\mathbf{v}|=1} \mu(\mathbf{v}). \quad (2.2)$$

DEFINITION 2.1. The quantity $\|\varphi\|$ determined by the formulas (2.1) and (2.2) is called the *norm* of a quadratic form φ in a Euclidean vector space V . If the norm $\|\varphi\|$ is finite, the form φ is said to be a *restricted quadratic form*.

THEOREM 2.1. If φ is a restricted quadratic form, then there is the estimate $|\varphi(\mathbf{v}, \mathbf{w})| \leq \|\varphi\| |\mathbf{v}| |\mathbf{w}|$ for the values of corresponding symmetric bilinear form.

PROOF. In order to calculate $\varphi(\mathbf{v}, \mathbf{w})$ we use the following equality, which, in essential, is a version of the recovery formula:

$$4\alpha \varphi(\mathbf{v}, \mathbf{w}) = \varphi(\mathbf{v} + \alpha \cdot \mathbf{w}) - \varphi(\mathbf{v} - \alpha \cdot \mathbf{w}). \quad (2.3)$$

From (2.3) we derive the following inequality for the quantity $4\alpha \varphi(\mathbf{v}, \mathbf{w})$:

$$4\alpha \varphi(\mathbf{v}, \mathbf{w}) \leq |\varphi(\mathbf{v} + \alpha \cdot \mathbf{w})| + |\varphi(\mathbf{v} - \alpha \cdot \mathbf{w})|. \quad (2.4)$$

Now let's apply the inequality $|\varphi(\mathbf{u})| \leq \|\varphi\| |\mathbf{u}|^2$ derived from (2.1) and (2.2) in order to estimate the right hand side of (2.4). This yields

$$4\alpha \varphi(\mathbf{v}, \mathbf{w}) \leq \|\varphi\| (|\mathbf{v} + \alpha \cdot \mathbf{w}|^2 + |\mathbf{v} - \alpha \cdot \mathbf{w}|^2). \quad (2.5)$$

Let's express the squares of moduli through the scalar products:

$$|\mathbf{v} \pm \alpha \cdot \mathbf{w}|^2 = |\mathbf{v}|^2 \pm 2\alpha (\mathbf{v} | \mathbf{w}) + \alpha^2 |\mathbf{w}|^2.$$

Then we can simplify the inequality (2.5) bringing it to the following one:

$$4\alpha \varphi(\mathbf{v}, \mathbf{w}) \leq 2\|\varphi\| (|\mathbf{v}|^2 + \alpha^2 |\mathbf{w}|^2).$$

Now let's transform the above inequality a little bit more:

$$f(\alpha) = \alpha^2 \|\varphi\| |\mathbf{w}|^2 - 2\alpha \varphi(\mathbf{v}, \mathbf{w}) + \|\varphi\| |\mathbf{v}|^2 \geq 0.$$

The numeric function $f(\alpha)$ of a numeric argument α is a polynomial of degree two in α . Let's find the minimum point $\alpha = \alpha_{\min}$ for this function by equating its derivative to zero: $f'(\alpha) = 0$. As a result we obtain

$$\alpha_{\min} = \frac{\varphi(\mathbf{v}, \mathbf{w})}{\|\varphi\| |\mathbf{w}|^2}.$$

Now let's write the inequality $f(\alpha_{\min}) \geq 0$ for the minimal value of this function. This yields the following inequality for the bilinear form φ :

$$\varphi(\mathbf{v}, \mathbf{w})^2 \leq \|\varphi\|^2 |\mathbf{v}|^2 |\mathbf{w}|^2.$$

Now it is easy to derive the required estimate for $|\varphi(\mathbf{v}, \mathbf{w})|$ by taking the square root of both sides of the above inequality. Note that a quite similar method was used when proving the Cauchy-Bunyakovsky-Schwarz inequality in the theorem 1.1. $\square$

THEOREM 2.2. *Any quadratic form φ in a finite-dimensional Euclidean vector space V is a restricted form.*

PROOF. Let's choose an orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in V and consider the expansion of a unit vector $\mathbf{v}$ in this basis. For the coordinates of $\mathbf{v}$ in this basis due to the formulas (1.12) we obtain

$$(v^1)^2 + \dots + (v^n)^2 = 1.$$

Hence, for the components of $\mathbf{v}$ we have $|v^i| \leq 1$. Let's express the quantity $\mu(\mathbf{v})$, which is defined by formula (2.1), through the coordinates of $\mathbf{v}$:

$$\mu(v) = |\varphi(v)| = \left| \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} v^i v^j \right|.$$

From $|v^i| \leq 1$ we derive the following estimate for the quantity $\mu(\mathbf{v})$:

$$\mu(v) \leq \sum_{i=1}^n \sum_{j=1}^n |\varphi_{ij}| < \infty, \quad (2.6)$$

Right hand side of (2.6) does not depend $\mathbf{v}$. Due to (2.2) this sum is an upper bound for the norm $\|\varphi\|$. Hence, $\|\varphi\| < \infty$. The theorem is proved. $\square$

THEOREM 2.3. *For any quadratic form φ in a finite-dimensional Euclidean vector space V the supremum in formula (2.2) is reached, i.e. there exists a vector $\mathbf{v} \neq \mathbf{0}$ such that $|\varphi(\mathbf{v})| = \|\varphi\| \|\mathbf{v}\|^2$.*

PROOF. From the course of mathematical analysis we know that the supremum of a numeric set is the limit of some converging sequence of numbers of this set (see [6]). This means that there is a sequence of unit vectors $\mathbf{v}(1), \dots, \mathbf{v}(n), \dots$ in V such that the norm $\|\varphi\|$ is expressed as the following limit:

$$\|\varphi\| = \lim_{s \rightarrow \infty} |\varphi(\mathbf{v}(s))|. \quad (2.7)$$

Let's choose an orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in $\mathbf{v}$ and let's expand each vector $\mathbf{v}(s)$ of the sequence in this basis. The equality

$$(v^1(s))^2 + \dots + (v^n(s))^2 = 1 \quad (2.8)$$

is derived from $|\mathbf{v}(s)| = 1$ due to the formulas (1.12). Now the equality (2.8) means that each specific coordinate $v^i(s)$ yields a restricted sequence of real numbers:

$$-1 \leq v^i(s) \leq 1.$$

From the course of mathematical analysis we know that in each restricted sequence of real numbers one can choose a converging subsequence. So, in the sequence

of unit vectors $\mathbf{v}(s)$ one can choose a subsequence of unit vectors whose first coordinates form a convergent sequence of numbers. Let's denote this subsequence again by $\mathbf{v}(s)$ and choose its subsequence with converging second coordinates. Repeating this choice n -times for each specific coordinate, we get a subsequence of unit vectors $\mathbf{v}(s_k)$ such that their coordinates all are the converging sequences of numbers. Let's consider the limits of these sequences:

$$v^i = \lim_{k \rightarrow \infty} v^i(s_k). \quad (2.9)$$

Denote by $\mathbf{v}$ the vector whose coordinates are determined by the limit values (2.9). Passing to the limit $s \rightarrow \infty$ in (2.8), we conclude that $\mathbf{v}$ is a unit vector: $|\mathbf{v}| = 1$.

Now let's calculate $|\varphi(\mathbf{v})|$ using the matrix of the quadratic form φ and the coordinates of $\mathbf{v}$ in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$:

$$|\varphi(\mathbf{v})| = \left| \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} v^i v^j \right| = \lim_{k \rightarrow \infty} \left| \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} v^i(s_k) v^j(s_k) \right|.$$

On the other hand, taking into account (2.7), for $|\varphi(\mathbf{v})|$ we get

$$|\varphi(\mathbf{v})| = \lim_{k \rightarrow \infty} |\varphi(v(s_k))| = \lim_{s \rightarrow \infty} |\varphi(v(s))| = \|\varphi\|. \quad (2.10)$$

Thus, for the unit vector $\mathbf{v}$ with coordinates (2.9) we get $|\varphi(\mathbf{v})| = \|\varphi\|$. Multiplying $\mathbf{v}$ by some number $\alpha \in \mathbb{R}$, we can remove the restriction $|\mathbf{v}| = 1$. Then the equality (2.10) will be written as $|\varphi(\mathbf{v})| = \|\varphi\| |\mathbf{v}|^2$. The theorem is proved. $\square$

THEOREM 2.4. *For any quadratic form φ in a finite-dimensional Euclidean vector space (V, g) there is an orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ such that the matrix of the form φ in this basis is a diagonal matrix.*

PROOF. The proof is by induction on the dimension of the space V . In the case $\dim V = 1$ the proposition of the theorem is obvious: any square 1×1 matrix is a diagonal matrix.

Suppose that the proposition of the theorem is valid for all quadratic forms in Euclidean spaces of the dimension less than n . Let $\dim V = n$ and let φ be a quadratic form in the Euclidean space (V, g) . Applying theorems 2.2 and 2.3, we find a unit vector $\mathbf{v} \in V$ such that $|\varphi(\mathbf{v})| = \|\varphi\|$. For the sake of certainty we assume that $\varphi(v) \geq 0$. Then we can remove the modulus sign: $\varphi(v) = \|\varphi\|$. In the case $\varphi(\mathbf{v}) < 0$ we replace the form φ by the opposite form $\tilde{\varphi} = -\varphi$ since two opposite forms diagonalize simultaneously.

Let's denote $U = \langle \mathbf{v} \rangle$ and consider the orthogonal complement $U_\perp$. The subspaces $U = \langle v \rangle$ and $U_\perp$ have zero intersection, their sum is a direct sum and $U \oplus U_\perp = V$ (see theorem 4.44 in Chapter IV). Let's take an arbitrary vector $\mathbf{w} \in U_\perp$ of the unit length and compose the vector $\mathbf{u}$ as follows:

$$\mathbf{u} = \cos(\alpha) \cdot \mathbf{v} + \sin(\alpha) \cdot \mathbf{w}.$$

Here α is a numeric parameter. It is easy to see that $\mathbf{u}$ is also a unit vector, this follows from the identity $\cos^2(\alpha) + \sin^2(\alpha) = 1$.

Let's calculate the value of the quadratic form φ on the vector $\mathbf{u}$ and treat it as a function of the numeric parameter α :

$$f(\alpha) = \varphi(\mathbf{u}) = \cos^2(\alpha) \varphi(\mathbf{v}) + 2 \sin(\alpha) \cos(\alpha) \varphi(\mathbf{v}, \mathbf{w}) + \sin^2(\alpha) \varphi(\mathbf{w}).$$

According to the choice of the vector v , we have the estimate $\varphi(\mathbf{u}) \leq \varphi(\mathbf{v})$, and for $\alpha = 0$, i.e. when $u = v$, we have the equality $\varphi(\mathbf{u}) = \varphi(\mathbf{v})$. Hence, $\alpha = 0$ is a maximum point for the function $f(\alpha)$. Let's calculate its derivative at the point $\alpha = 0$ and equate it to zero. This yields

$$f'(0) = 2 \varphi(\mathbf{v}, \mathbf{w}) = 0. \quad (2.11)$$

Hence, $\varphi(\mathbf{v}, \mathbf{w}) = 0$ for all vectors $\mathbf{w} \in U_\perp$. Let's apply the inductive hypothesis to the subspace $U_\perp$ whose dimension is less by 1 than the dimension of the space V . Therefore, we can find an orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$ in the subspace $U_\perp$ such that the matrix of the form φ is diagonal in this basis: $\varphi(\mathbf{e}_i, \mathbf{e}_j) = 0$ for $i \neq j$. Let's complete the basis $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$ with the vector $\mathbf{e}_n = \mathbf{v}$. The complementary vector $\mathbf{e}_n$ is a vector of unit length. It is orthogonal to the vectors $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$. Therefore, the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ is an orthonormal basis in V . The matrix of the form φ is diagonal in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. This fact is immediate from (2.11). The theorem is proved. $\square$

The theorem 2.4 is known as the theorem on simultaneous diagonalization of a pair quadratic form φ and g . For this purpose one of them should be positive. Then the positive form g defines the structure of a Euclidean space in V and then one can apply the theorem 2.4. Orthonormality of the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ means that the matrix of g is diagonal in this basis (it is the unit matrix). The matrix of φ is also diagonal as stated in the theorem 2.4.

§ 3. Selfadjoint operators. The theorem on the spectrum and the basis of eigenvectors for a selfadjoint operator.

DEFINITION 3.1. A linear operator $f: V \rightarrow V$ in a Euclidean vector space V is called a *symmetric operator* or a *selfadjoint operator* if for any two vectors $\mathbf{v}, \mathbf{w} \in V$ the following equality is fulfilled: $(\mathbf{v} | f(\mathbf{w})) = (f(\mathbf{v}) | \mathbf{w})$.

DEFINITION 3.2. A linear operator $h: V \rightarrow V$ in a Euclidean vector space V is called an *adjoint operator* to the operator $f: V \rightarrow V$ if for any two vectors $\mathbf{v}, \mathbf{w} \in V$ the following equality is fulfilled: $(\mathbf{v} | f(\mathbf{w})) = (h(\mathbf{v}) | \mathbf{w})$. The adjoint operator is denoted as follows: $h = f^+$.

In S 4 of Chapter III we have introduced the concept of conjugate mapping. There we have shown that any linear mapping $f: V \rightarrow W$ possesses the conjugate mapping $f^*: W^* \rightarrow V^*$. For a linear operator $f: V \rightarrow V$ the conjugate mapping f^* is a linear operator in dual space V^* . It is related to f by means of the equality

$$\langle f^*(\mathbf{u}) | \mathbf{v} \rangle = \langle \mathbf{u} | f(\mathbf{v}) \rangle, \quad (3.1)$$

which is fulfilled for all $\mathbf{u} \in V^*$ and for all $\mathbf{v} \in V$.

The structure of a Euclidean vector space in V is determined by a positive quadratic form g . Like every quadratic form, the form g possesses the associated mapping $a_g: V \rightarrow V^*$ (see § 2 in Chapter III) such that

$$\langle a_g(\mathbf{v}) | \mathbf{w} \rangle = g(\mathbf{v}, \mathbf{w}) = (\mathbf{v} | \mathbf{w}). \quad (3.2)$$

In the case of finite-dimensional space V and positive form g the associated mapping a_g is bijective. Therefore, for any linear operator $f: V \rightarrow V$ we can define the composition $h = a_g^{-1} \circ f^* \circ a_g$. Then from (3.1) and (3.2) we derive

$$\begin{aligned} (h(\mathbf{v}) | \mathbf{w}) &= \langle a_g \circ h(\mathbf{v}) | \mathbf{w} \rangle = \\ &= \langle f^* \circ a_g(\mathbf{v}) | \mathbf{w} \rangle = \langle a_g(\mathbf{v}) | f(\mathbf{w}) \rangle = (\mathbf{v} | f(\mathbf{w})). \end{aligned} \quad (3.3)$$

Comparing (3.3) with the definition 3.2, we can formulate the following theorem.

THEOREM 3.1. *For any operator f in a finite-dimensional Euclidean space (V, g) there is the unique adjoint operator $f^+ = a_g^{-1} \circ f^* \circ a_g$.*

PROOF. The existence of an adjoint operator is already derived from the formula $f^+ = a_g^{-1} \circ f^* \circ a_g$ and the equality (3.3). Let's prove its uniqueness. Assume that h is another operator satisfying the definition 3.2. Then for the difference $r = h - f^+$ we derive the relationship

$$\begin{aligned} (r(\mathbf{v}) | \mathbf{w}) &= (h(\mathbf{v}) | \mathbf{w}) - (f^+(\mathbf{v}) | \mathbf{w}) = \\ &= (\mathbf{v} | f(\mathbf{w})) - (\mathbf{v} | f(\mathbf{w})) = 0. \end{aligned} \quad (3.4)$$

Since $\mathbf{w}$ in (3.4) is an arbitrary vector, we conclude that $r(\mathbf{v}) \in \text{Ker } g$. However, $\text{Ker } g = \{\mathbf{0}\}$ for a positive quadratic form g , hence, $h(\mathbf{v}) = \mathbf{0}$ for any $\mathbf{v} \in V$. This means that $h = 0$. Thus, we have proved that the adjoint operator f^+ for f is unique. This completes the proof of the theorem. $\square$

COROLLARY. *The passage from f to f^+ is an operator in the space of endomorphisms $\text{End}(V)$ of a finite-dimensional Euclidean vector space (V, g) . This operator possesses the following properties:*

$$\begin{aligned} (f + h)^+ &= f^+ + h^+, & (\alpha \cdot f)^+ &= \alpha \cdot f^+, \\ (f \circ h)^+ &= h^+ \circ f^+, & (f^+)^+ &= f. \end{aligned}$$

Relying upon the existence and the uniqueness of the adjoint operator f^+ for any operator $f \in \text{End}(V)$, we can derive all the above relationships immediately from the definition 3.2. The relationship $f^+ = a_g^{-1} \circ f^* \circ a_g$ can be expressed in the form of the following commutative diagram:

$$\begin{array}{ccc} V & \xrightarrow{f^+} & V \\ a_g \downarrow & & \downarrow a_g \\ V^* & \xrightarrow{f^*} & V^* \end{array}$$

Comparing the definitions 3.1 and 3.2, now we see that a selfadjoint operator f is an operator which is adjoint to itself: $f^+ = f$.

Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis in a finite-dimensional Euclidean space (V, g) and let $h^1, \dots, h^n$ be the corresponding dual basis composed by coordinate functionals. For any vector $\mathbf{v} \in V$ we have the following expansion, which follows from the definition of coordinate functionals (see § 13 in Chapter III):

$$\mathbf{v} = h^1(\mathbf{v}) \cdot \mathbf{e}_1 + \dots + h^n(\mathbf{v}) \cdot \mathbf{e}_n.$$

Let's apply this expansion in order to calculate the matrix of the associated mapping a_g . For this purpose we need to apply a_g one by one to all basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ and expand the results in the dual basis in V^* . Let's consider the value of the functional $a_g(\mathbf{e}_i)$ on an arbitrary vector $\mathbf{v}$ of the space V :

$$\begin{aligned} a_g(\mathbf{e}_i)(\mathbf{v}) &= \langle a_g(\mathbf{e}_i) | \mathbf{v} \rangle = g(\mathbf{e}_i, \mathbf{v}) = \\ &= g(\mathbf{e}_i, h^1(\mathbf{v}) \cdot \mathbf{e}_1 + \dots + h^n(\mathbf{v}) \cdot \mathbf{e}_n) = \sum_{j=1}^n g_{ij} h^j(\mathbf{v}). \end{aligned}$$

Since $\mathbf{v} \in V$ is an arbitrary vector, we conclude that the matrix of the associated mapping a_g in two bases $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $h^1, \dots, h^n$ coincides with the matrix $g_{ij} = g(\mathbf{e}_i, \mathbf{e}_j)$ of the quadratic form g in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. The matrix g_{ij} is non-degenerate (see theorem 1.2 or Sylvester's criterion in § 4 of Chapter IV). Let's denote by g^{ij} the components of the matrix inverse to g_{ij} . The matrix g^{ij} is the matrix of the inverse mapping a_g^{-1} , i.e. we have:

$$a_g(\mathbf{e}_i) = \sum_{j=1}^n g_{ij} h^j, \quad a_g^{-1}(h^j) = \sum_{i=1}^n g^{ij} \mathbf{e}_i. \quad (3.5)$$

The matrix inverse to a symmetric matrix is again a symmetric matrix (this fact is well-known from general algebra). Therefore $g^{ij} = g^{ji}$.

Remember that we have already calculated the matrix of the conjugate mapping f^* (see formula (4.2) and theorem 4.3 in Chapter III). When applied to our present case the results of Chapter III mean that the matrix of the operator $f^*: V^* \rightarrow V^*$ in the basis of coordinate functionals $h^1, \dots, h^n$ coincides with the matrix of the initial operator f in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. Let's combine this fact with (3.5) and let's use the formula $f^+ = a_g^{-1} \circ f^* \circ a_g$ from the theorem 3.1. Then for the matrix of F^+ of the adjoint operator f^+ we obtain:

$$(F^+)_j^i = \sum_{k=1}^n \sum_{q=1}^n g^{iq} F_q^k g_{kj}. \quad (3.6)$$

In matrix form the formula (3.6) is written as $F^+ = G^{-1} F^{\text{tr}} G$, where G is the Gram matrix of that basis in which the matrices of f and f^+ are calculated. The formula (3.6) simplifies substantially for orthonormal bases. Here the passage to the adjoint operator means the transposition of its matrix. The matrix of a selfadjoint operator in an orthonormal basis is symmetric. For this reason selfadjoint operators are often called symmetric operators.

Let $f: V \rightarrow V$ be a selfadjoint operator in a Euclidean space V . Each such operator produces the quadratic φ_f according to the formula

$$\varphi_f(\mathbf{v}) = (\mathbf{v} | f(\mathbf{v})). \quad (3.7)$$

Conversely, assume that we have a quadratic form φ in a finite-dimensional Euclidean space (V, g) . The form φ determines the associated mapping a_φ (see definition 2.5 in Chapter IV). This mapping satisfies the relationship

$$\langle a_\varphi(\mathbf{v}) | \mathbf{w} \rangle = \varphi(\mathbf{v}, \mathbf{w}) \quad (3.8)$$

for any two vectors $\mathbf{v}, \mathbf{w} \in V$. The positive quadratic form g defining the structure of Euclidean space in V has also its own associated mapping a_g . The mapping a_g is bijective since g is non-degenerate (see theorem 4.2 in Chapter IV). Therefore, we can consider the composition of a_g^{-1} and a_φ :

$$f_\varphi = a_g^{-1} \circ a_\varphi. \quad (3.9)$$

This composition (3.9) is an operator in V . It is called the *associated operator* of the form φ in a Euclidean space. Since a_g is bijective, we can write (3.2) as $\langle \mathbf{u} | \mathbf{w} \rangle = \langle a_g^{-1}(\mathbf{u}) | \mathbf{w} \rangle$. Combining this equality with (3.8), we find

$$(f_\varphi(\mathbf{v}) | \mathbf{w}) = (a_g^{-1}(a_\varphi(\mathbf{v})) | \mathbf{w}) = \langle a_\varphi(\mathbf{v}) | \mathbf{w} \rangle = \varphi(\mathbf{v}, \mathbf{w}). \quad (3.10)$$

Now, using the symmetry of the form $\varphi(\mathbf{v}, \mathbf{w})$ in (3.10), we write

$$(f_\varphi(\mathbf{v}) | \mathbf{w}) = \varphi(\mathbf{v}, \mathbf{w}) = \varphi(\mathbf{w}, \mathbf{v}) = (f_\varphi(\mathbf{w}) | \mathbf{v}) = (\mathbf{v} | f_\varphi(\mathbf{w})). \quad (3.11)$$

The relationship (3.11), which is an identity for all $\mathbf{v}, \mathbf{w} \in V$, means that f_φ is a selfadjoint operator (see definition 2.1).

The formula (3.7) associates each selfadjoint operator f with the quadratic form φ_f , while the formula (3.9) associates each quadratic form φ with the selfadjoint operator f_φ . These two associations are one to one and are inverse to each other. Indeed, let's apply the formula (3.7) to the operator (3.9) and use (3.10):

$$\varphi_f(\mathbf{v}) = (\mathbf{v} | f_\varphi(\mathbf{v})) = \varphi(\mathbf{v}, \mathbf{v}) = \varphi(\mathbf{v}).$$

Now, conversely, let's construct the operator $h = f_\varphi$ for the quadratic form $\varphi = \varphi_f$. For the operator h and for two arbitrary vectors $\mathbf{v}, \mathbf{w} \in V$ from (3.10) we derive

$$(h(\mathbf{v}) | \mathbf{w}) = \varphi_f(\mathbf{v}, \mathbf{w}) = (\mathbf{v} | f(\mathbf{w})) = (f(\mathbf{v}) | \mathbf{w}).$$

Since $\mathbf{w} \in V$ is an arbitrary vector and since the form g determining the scalar product in V is non-degenerate, from the above equality we get $h(\mathbf{v}) = f(\mathbf{v})$.

Thus, from what was said above we conclude that defining a selfadjoint operator in a finite-dimensional Euclidean space is equivalent to defining a quadratic form in this space. Therefore, we can apply the theorem 2.4 for describing selfadjoint operators in a finite-dimensional case.

THEOREM 3.2. *All eigenvalues of a selfadjoint operator f in a finite-dimensional Euclidean space V are real numbers and there is an orthonormal basis composed by eigenvectors of such operator.*

PROOF. For the selfadjoint operator f in V we consider the symmetric bilinear form $\varphi_f(\mathbf{v}, \mathbf{w})$ determined by the quadratic form (3.7). Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be an orthonormal basis in which the matrix of the form φ_f is diagonal. Then from the formula (3.7) we derive the following equalities:

$$\varphi_f(\mathbf{e}_i, \mathbf{e}_j) = (\mathbf{e}_i | f(\mathbf{e}_j)) = \sum_{k=1}^n F_j^k g_{ik} = F_k^i. \quad (3.12)$$

As we see in (3.12), the matrices of the operator f and of the form φ_f in such basis do coincide. This proves the proposition of the theorem. $\square$

The theorem 3.2 is known as the theorem on the spectrum and the basis of eigenvectors of a selfadjoint operator. The main result of this theorem is the diagonalizability of selfadjoint operators in a finite-dimensional Euclidean space. The characteristic polynomial of a selfadjoint operator is factorized into the product of linear terms in $\mathbb{R}$. Its eigenspaces coincide with the corresponding root subspaces, the sum of all its eigenspaces coincides with the space V :

$$V = V_{\lambda_1} \oplus \dots \oplus V_{\lambda_s}. \quad (3.13)$$

THEOREM 3.3. *Any two eigenvectors of a selfadjoint operator corresponding to different eigenvalues are orthogonal to each other.*

PROOF. Let f be a selfadjoint operator in a Euclidean space and let $\lambda \neq \mu$ be its eigenvalues. Let's consider the corresponding eigenvectors $\mathbf{a}$ and $\mathbf{b}$:

$$f(\mathbf{a}) = \lambda \cdot \mathbf{a}, \quad f(\mathbf{b}) = \mu \cdot \mathbf{b}.$$

Then for these two eigenvectors $\mathbf{a}$ and $\mathbf{b}$ we derive:

$$\lambda(\mathbf{a} | \mathbf{b}) = (f(\mathbf{a}) | \mathbf{b}) = (\mathbf{a} | f(\mathbf{b})) = \mu(\mathbf{a} | \mathbf{b}).$$

Hence, $(\lambda - \mu)(\mathbf{a} | \mathbf{b}) = 0$. But we know that $\lambda - \mu \neq 0$. Therefore, $(\mathbf{a} | \mathbf{b}) = 0$. The theorem is proved. $\square$

Assume that the kernel of selfadjoint operator f is nontrivial: $\text{Ker } f \neq \{\mathbf{0}\}$. Then $\lambda_1 = 0$ in (3.13) is one of the eigenvalues of the operator f and we have

$$\text{Ker } f = V_{\lambda_1}, \quad \text{Im } f = V_{\lambda_2} \oplus \dots \oplus V_{\lambda_s}.$$

This means that the kernel and the image of a selfadjoint operator are orthogonal to each other and their sum coincides with V :

$$V = \text{Ker } f \oplus \text{Im } f. \quad (3.14)$$

§ 4. Isometries and orthogonal operators.

DEFINITION 4.1. A linear mapping $f: V \rightarrow W$ from one Euclidean vector space (V, g) to another Euclidean vector space (W, h) is called an *isometry* if

$$(f(\mathbf{x}) | f(\mathbf{y})) = (\mathbf{x} | \mathbf{y}) \quad (4.1)$$

for all $\mathbf{x}, \mathbf{y} \in V$, i. e. if it preserves the scalar product of vectors.

From (4.1) we easily derive $|f(\mathbf{x})| = |\mathbf{x}|$, therefore, $f(\mathbf{x}) = \mathbf{0}$ implies $|\mathbf{x}| = 0$ and $\mathbf{x} = \mathbf{0}$. This means that the kernel of an isometry is always trivial $\text{Ker } f = \{\mathbf{0}\}$, i. e. any isometry is an injective mapping. Due to the recovery formula for quadratic forms (see formula (1.6) in Chapter IV) in order to verify that $f: V \rightarrow W$ is an isometry it is sufficient to verify that it preserves the norm of vectors, i. e. $|f(\mathbf{x})| = |\mathbf{x}|$ for all vectors $\mathbf{x} \in V$.

THEOREM 4.1. *The composition of isometries is again an isometry.*

PROOF. Assume that the mappings $h: U \rightarrow V$ and $f: V \rightarrow W$ both are isometries. Hence, $|h(\mathbf{u})| = |\mathbf{u}|$ for all $\mathbf{u} \in U$ and $|f(\mathbf{v})| = |\mathbf{v}|$ for all $\mathbf{v} \in V$. Then

$$|f \circ h(\mathbf{u})| = |f(h(\mathbf{u}))| = |h(\mathbf{u})| = |\mathbf{u}|$$

for all $\mathbf{u} \in U$. This equality means that the mapping $f \circ h$ is an isometry. The theorem is proved. $\square$

DEFINITION 4.2. A bijective isometry $f: V \rightarrow W$ is called an *isomorphism of Euclidean vector spaces*.

THEOREM 4.2. *Isomorphisms of Euclidean vector spaces possess the following three properties:*

- (1) *the identical mapping id_V is an isomorphism;*
- (2) *the composition of isomorphisms is an isomorphism;*
- (3) *the mapping inverse to an isomorphism is an isomorphism.*

The proof of this theorem is very easy if we use the above theorem 4.1 and the theorem 8.1 of Chapter I.

DEFINITION 4.3. Two Euclidean vector spaces V and W are called *isomorphic* if there is an isomorphism $f: V \rightarrow W$ relating them.

Let's consider the arithmetic vector space $\mathbb{R}^n$ composed by column vectors of the height n . The addition of such vectors and the multiplication of them by real numbers are performed as the operations with their components (see formulas (2.1) in Chapter I). Let's define a quadratic form $g(x)$ in $\mathbb{R}^n$ by setting

$$g(\mathbf{x}) = (x^1)^2 + \dots + (x^n)^2 = \sum_{i=1}^n (x^i)^2. \quad (4.2)$$

The form (4.2) yields the *standard scalar product* and, hence, defines the *standard structure of a Euclidean space* in $\mathbb{R}^n$.

THEOREM 4.3. *Any n -dimensional Euclidean vector space V is isomorphic to the space $\mathbb{R}^n$ with the standard scalar product (4.2).*

In order to prove this theorem it is sufficient to choose the orthonormal basis in V and consider the mapping ψ that associates a vector $\mathbf{v} \in V$ with column vector of its coordinates (see formula (5.4) in Chapter I).

DEFINITION 4.4. An operator f in a Euclidean vector space V is called an *orthogonal operator* if it is bijective and defines an isometry $f: V \rightarrow V$.

Due to the theorem 4.2 the orthogonal operators form a group which is called the *orthogonal group* of a Euclidean space V and is denoted by $O(V)$. The group $O(V)$ is obviously a subgroup in the group of automorphisms $\text{Aut}(V)$. In the case $V = \mathbb{R}^n$ the orthogonal group determined by the standard scalar product in $\mathbb{R}^n$ is denoted by $O(n, \mathbb{R})$.

Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be an orthonormal basis in a Euclidean space V and let f be an orthogonal operator. Then from (4.1) we derive

$$(f(\mathbf{e}_i) | f(\mathbf{e}_j)) = (\mathbf{e}_i | \mathbf{e}_j).$$

For the matrix of the operator f in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ this relationship yields:

$$\sum_{k=1}^n F_i^k F_j^k = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j, \end{cases} \quad (4.3)$$

When written in the matrix form, the formula (4.3) means that

$$F^{\text{tr}} F = 1, \quad F^{-1} = F^{\text{tr}}. \quad (4.4)$$

The relationships (4.4) are identical to the relationships (1.13). Matrices that satisfy such relationships, as we already know, are called *orthogonal matrices*. As a corollary of this fact we can formulate the following theorem.

THEOREM 4.4. *An orthogonal operator f in an orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of a Euclidean vector space V is given by an orthogonal matrix.*

As we have noted in § 1, the determinant of an orthogonal matrix can be equal to 1 or to -1 . The orthogonal operators in V with determinant 1 form a group which is called the *special orthogonal group* of a Euclidean vector space V . This group is denoted by $SO(V)$. If $V = \mathbb{R}^n$, this group is denoted by $SO(n, \mathbb{R})$.

The operators $f \in SO(V)$ in two-dimensional case $\dim V = 2$ are most simple ones. If $\mathbf{e}_1, \mathbf{e}_2$ is an orthonormal basis in V , then from (4.3) and $\det F = 1$ we easily find the form of an orthogonal matrix F :

$$F = \begin{vmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{vmatrix} \quad (4.5)$$

A matrix F of the form (4.5) is called a *matrix of two-dimensional rotation*, while the numeric parameter φ is interpreted as the angle of rotation.

Let's consider orthogonal operators $f \in \text{SO}(V)$ in the case $\dim V = 3$. Let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ be an orthonormal basis in V . A matrix of the form

$$F = \begin{vmatrix} \cos(\varphi) & -\sin(\varphi) & 0 \\ \sin(\varphi) & \cos(\varphi) & 0 \\ 0 & 0 & 1 \end{vmatrix} \quad (4.6)$$

is an orthogonal matrix with determinant 1. The operator f associated with the matrix (4.6) is called the operator of *rotation* about the vector $\mathbf{e}_3$ by the angle φ .

THEOREM 4.5. *In a three-dimensional Euclidean vector space V any orthogonal operator f with determinant 1 has an eigenvalue $\lambda = 1$.*

PROOF. Let's consider the characteristic polynomial of the operator f . This is the polynomial of degree 3 in λ with real coefficients:

$$P(\lambda) = -\lambda^3 + F_1 \lambda^2 - F_2 \lambda + F_3, \text{ where } F_3 = \det f = 1.$$

Remember that the values of a polynomial of odd degree for large positive λ and for large negative λ differ in sign:

$$\lim_{\lambda \rightarrow -\infty} P(\lambda) = +\infty, \quad \lim_{\lambda \rightarrow +\infty} P(\lambda) = -\infty.$$

Therefore the equation of the odd degree $P(\lambda) = 0$ with real coefficients has at least one real root $\lambda = \lambda_1$. This root is an eigenvalue of the operator f .

Let $\mathbf{e}_1 \neq \mathbf{0}$ be an eigenvector of f corresponding to the eigenvalue λ_1 . Then, applying the isometry condition $|\mathbf{v}| = |f(\mathbf{v})|$ to the vector $\mathbf{v} = \mathbf{e}_1$, we get

$$|\mathbf{e}_1| = |f(\mathbf{e}_1)| = |\lambda_1 \cdot \mathbf{e}_1| = |\lambda_1| |\mathbf{e}_1|.$$

Hence, we find that $|\lambda_1| = 1$. This means that $\lambda_1 = 1$ or $\lambda_1 = -1$. In the case $\lambda_1 = 1$ the proposition of the theorem is valid. Therefore, we consider the case $\lambda_1 = -1$. Let's separate the linear factor $(\lambda + 1)$ in characteristic polynomial:

$$P(\lambda) = -\lambda^3 + F_1 \lambda^2 - F_2 \lambda + 1 = -(\lambda + 1)(\lambda^2 - \Phi_1 \lambda - 1).$$

Then $F_1 = \Phi_1 - 1$ and $F_2 = -1 - \Phi_1$. In order to the remaining roots of the polynomial $P(\lambda)$ we consider the following quadratic equation:

$$\lambda^2 - \Phi_1 \lambda - 1 = 0.$$

This equation always has two real roots λ_2 and λ_3 since its discriminant is positive: $D = (\Phi_1)^2 + 4 > 0$. Due to the Viet theorem we have $\lambda_2 \lambda_3 = -1$. Due to the same reasons as above in the case of λ_1 , for λ_2 and λ_3 we get $|\lambda_2| = |\lambda_3| = 1$. Hence, one of these two real numbers is equal to 1 and the other is equal to -1 . Thus, we have proved that the number $\lambda = 1$ is among the eigenvalues of the operator f . The theorem is proved. $\square$

THEOREM 4.5. *In a three-dimensional Euclidean vector space V for any orthogonal operator f with determinant 1 there is an orthonormal basis in which the matrix of f has the form (4.6).*

PROOF. Under the assumptions of theorem 4.5 the operator f has an eigenvalue $\lambda_1 = 1$. Let $\mathbf{e}_1 \neq \mathbf{0}$ be an eigenvector of this operator associated with the eigenvalue $\lambda_1 = 1$. Let's denote by U the span of the eigenvector $\mathbf{e}_1$ and consider its orthogonal complement $U_\perp$. This is the two-dimensional subspace in the three-dimensional space V . This subspace is invariant under the action of f . Indeed, from $\mathbf{x} \in U_\perp$ we derive $(\mathbf{x} | \mathbf{e}_1) = 0$. Let's write the isometry condition (4.1) for the vectors $\mathbf{x}$ and $\mathbf{y} = \mathbf{e}_1$:

$$0 = (\mathbf{x} | \mathbf{e}_1) = (f(\mathbf{x}) | f(\mathbf{e}_1)) = \lambda_1 (f(\mathbf{x}) | \mathbf{e}_1).$$

Since $\lambda_1 = 1$, we get $(f(\mathbf{x}) | \mathbf{e}_1) = 0$. Hence, $f(\mathbf{x}) \in U_\perp$, which proves the invariance of the subspace $U_\perp$.

Let's consider the restriction of the operator f to the invariant subspace $U_\perp$. This restriction is an orthogonal operator in two-dimensional space $U_\perp$, its determinant being equal to 1. Therefore, in some orthogonal basis $\mathbf{e}_2, \mathbf{e}_3$ of $U_\perp$ the matrix of the restricted operator has the form (4.5).

Remember that $\mathbf{e}_1$ is perpendicular to $\mathbf{e}_2$ and $\mathbf{e}_3$. It can be normalized to the unit length. Then three vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ form an orthonormal basis in three-dimensional space V and the matrix of f in this basis has the form (4.6). The theorem is proved. $\square$

The result of this theorem is that any orthogonal operator f with determinant 1 in a three-dimensional Euclidean vector space V is an operator of rotation. The eigenvector $\mathbf{e}_1$ associated with the eigenvalue $\lambda_1 = 1$ determines the axis of rotation, while the real parameter φ in the matrix (4.6) determines the angle of such rotation.

CHAPTER VI

AFFINE SPACES.

§ 1. Points and parallel translations. Affine spaces.

Let M be an arbitrary set. A *transformation* of the set M is a bijective mapping $p: M \rightarrow M$ of the set M onto itself.

DEFINITION 1.1. Let V be a linear vector space. We say that an *action* of V on a set M is defined if each vector $\mathbf{v} \in V$ is associated with some transformation $p_{\mathbf{v}}$ of the set M and the following conditions are fulfilled:

- (1) $p_{\mathbf{0}} = \text{id}_M$;
- (2) $p_{\mathbf{v}+\mathbf{w}} = p_{\mathbf{v}} \circ p_{\mathbf{w}}$ for all $\mathbf{v}, \mathbf{w} \in V$.

From the properties (1) and (2) of an action of a space V on a set M one can easily derive the following two properties of such action:

- (3) $p_{-\mathbf{v}} = p_{\mathbf{v}}^{-1}$ for all $\mathbf{v} \in V$;
- (4) $p_{\mathbf{v}} \circ p_{\mathbf{w}} = p_{\mathbf{w}} \circ p_{\mathbf{v}}$ for all $\mathbf{v}, \mathbf{w} \in V$.

DEFINITION 1.2. An action of a vector space V on a set M is called a *transitive action* if for any two elements $A, B \in M$ there is a vector $\mathbf{v} \in V$ such that $p_{\mathbf{v}}(A) = B$, i.e. the transformation $p_{\mathbf{v}}$ takes A to B .

DEFINITION 1.3. An action of a vector space V on a set M is called a *free action* if for any element $A \in M$ the equality $p_{\mathbf{v}}(A) = A$ implies $\mathbf{v} = \mathbf{0}$.

DEFINITION 1.4. A set M is called an *affine space* over the field $\mathbb{K}$ if there is a free transitive action of some linear vector space V over the field $\mathbb{K}$ on M .

Due to this definition any affine space M is associated with some linear vector space V . Therefore an affine space M is often denoted as a pair (M, V) .

Elements of an affine space are used to be called *points*. We shall denote them by capital letters A, B, C , etc. An affine space itself is sometimes called a *point space*. A transformation $p_{\mathbf{v}}$ given by a vector $\mathbf{v} \in V$ is called a *parallel translation* in an affine space M .

Let U be a subspace in V . Let's choose a point $A \in M$ and then let's define a subset $L \subset M$ in the following way:

$$L = \{B \in M : \exists \mathbf{u} ((\mathbf{u} \in U) \ \& \ (B = p_{\mathbf{u}}(A)))\}. \quad (1.1)$$

A subset L of M determined according to (1.1) is called a *linear submanifold* of an affine space M . Thereby the subspace $U \subset V$ is called the *directing subspace* of a linear submanifold L . The dimension of the directing subspace in (1.1) is taken for the *dimension* of a linear submanifold L . One-dimensional linear submanifolds are called *straight lines*; two-dimensional submanifolds are called *planes*. If the

dimension of U is less by one than the dimension of V , i.e. if $\dim(V/U) = 1$, then the corresponding linear submanifold L is called a *hyperplane*. Linear submanifolds of other intermediate dimensions have no special titles.

Let $U = \langle a \rangle$ be a one-dimensional subspace in V . Then any vector $\mathbf{u} \in U$ is presented as $\mathbf{u} = t \cdot \mathbf{a}$, where $t \in \mathbb{K}$. Upon choosing a point $A \in M$ the subspace U determines the straight line in M passing through the point A . An arbitrary point $A(t)$ of this straight is given by the formula:

$$A(t) = p_{t \cdot \mathbf{a}}(A). \quad (1.2)$$

The formula (1.2) is known as a *parametric equation* of a straight line in an affine space, the vector $\mathbf{a}$ is called a *directing vector*, while $t \in \mathbb{K}$ is a *parameter*.

If $\mathbb{K} = \mathbb{R}$, we can consider the set of points on the straight line (1.2) corresponding to the values of t taken from the interval $[0, 1] \subset \mathbb{R}$. Such set is called a *segment* of a straight line. The points $A = A(0)$ and $B = A(1)$ are ending points of this segment. One can choose a direction on the segment AB by saying that one of the ending points is the *beginning* of the segment and the other is the *end* of the segment. A segment AB with a fixed direction on it is called a *directed segment* or an *arrowhead segment*. Two arrowhead segments $\overrightarrow{AB}$ and $\overrightarrow{BA}$ are assumed to be distinct¹.

Let A and B be two points of an affine space M . Due to the transitivity of the action of V on M there exists a vector $\mathbf{v} \in V$ that defines the parallel translation $p_{\mathbf{v}}$ taking the point A to the point B : $p_{\mathbf{v}}(A) = B$. Let's prove that such parallel translation is unique. If $p_{\mathbf{w}}$ is another parallel translation such that $p_{\mathbf{w}}(A) = B$, then for the parallel translation $p_{\mathbf{w}-\mathbf{v}}$ we have

$$p_{\mathbf{w}-\mathbf{v}}(A) = p_{-\mathbf{v}} \circ p_{\mathbf{w}}(A) = p_{\mathbf{v}}^{-1}(p_{\mathbf{w}}(A)) = p_{\mathbf{v}}^{-1}(B) = A.$$

Since V acts freely on M (see definition 1.3), we have $\mathbf{w} - \mathbf{v} = \mathbf{0}$. Hence, $\mathbf{w} = \mathbf{v}$, this proves the uniqueness of the vector $\mathbf{v}$ determined by the condition $p_{\mathbf{v}}(A) = B$.

The above fact appears to be very useful: if we have an affine space (M, V) , then vectors of V can be represented by arrowhead segments in M . Each pair of points $A, B \in M$ specifies the unique vector $\mathbf{a} \in V$ such that $p_{\mathbf{a}}(A) = B$. This vector can be used as a directing vector of the straight line (1.2) passing through the points A and B . The arrowhead segment with the beginning at the point A and with the end at the point B is called the *geometric representation* of the vector $\mathbf{a}$. It is denoted $\overrightarrow{AB}$.

A vector $\mathbf{a}$ is uniquely determined by its geometric representation $\overrightarrow{AB}$. However, a vector $\mathbf{a}$ can have several geometric representations. Indeed, if we choose a point $C \neq A$, we can determine the point $D = p_{\mathbf{a}}(C)$ and then we can construct the geometric representation $\overrightarrow{CD}$ for the vector $\mathbf{a}$. The points A and C specify a parallel translation $p_{\mathbf{b}}$ such that $p_{\mathbf{b}}(A) = C$. Using the property (4) of parallel translations, it is easy to find that the parallel translation $p_{\mathbf{b}}$ maps the segment AB to the segment CD . So we conclude: various geometric representations of a vector $\mathbf{a}$ are related to each other by means of parallel translations. Note that $p_{\mathbf{a}+\mathbf{b}}(A) = D$. Therefore, $\overrightarrow{AD}$ is a geometric realization of the vector $\mathbf{a} + \mathbf{b}$. From

¹ If $\mathbb{K} \neq \mathbb{R}$, an arrowhead segment $\overrightarrow{AB}$ is assumed to be consisting on two points A and B only, it has no interior at all.

this fact we easily derive the well-known rules for vector addition: the triangle rule $\overrightarrow{AC} + \overrightarrow{CD} = \overrightarrow{AD}$ and the parallelogram rule $\overrightarrow{AB} + \overrightarrow{AC} = \overrightarrow{AD}$.

Let O be some fixed point of an affine space M . Let's call it the *origin*. Then any point $A \in M$ specifies the arrowhead $\overrightarrow{OA}$ which is identified with the unique vector $\mathbf{r} \in V$ by means of the equality $p_{\mathbf{r}}(O) = A$. This vector $\mathbf{r} = \mathbf{r}_A$ is called the *radius-vector* of the point A . If the space V is finite-dimensional, then we can choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ and then can expand the radius-vectors of all points $A \in M$ in this basis.

DEFINITION 1.5. A *frame* or a *coordinate system* in an affine space M is a pair consisting of a point $O \in M$ and a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ in V . The coordinates of the radius-vector $\mathbf{r}_A = \overrightarrow{OA}$ in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ are called the *coordinates* of a point A in the coordinate system $O, \mathbf{e}_1, \dots, \mathbf{e}_n$.

Coordinate systems in affine spaces play the same role as bases in linear vector spaces. Let $O, \mathbf{e}_1, \dots, \mathbf{e}_n$ and $O', \tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ be two coordinate systems in an affine space M . The relation of the bases $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ is given by the direct and inverse transition matrices S and T . The points O and O' determine the arrowhead segment $\overrightarrow{OO'}$ and the opposite arrowhead segment $\overrightarrow{O'O}$. They are associated with two vectors $\boldsymbol{\rho}, \tilde{\boldsymbol{\rho}} \in V$:

$$\boldsymbol{\rho} = \overrightarrow{OO'}, \quad \tilde{\boldsymbol{\rho}} = \overrightarrow{O'O}$$

Let's expand $\boldsymbol{\rho}$ in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $\tilde{\boldsymbol{\rho}}$ in the basis $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$:

$$\begin{aligned} \boldsymbol{\rho} &= \rho^1 \cdot \mathbf{e}_1 + \dots + \rho^n \cdot \mathbf{e}_n, \\ \tilde{\boldsymbol{\rho}} &= \tilde{\rho}^1 \cdot \tilde{\mathbf{e}}_1 + \dots + \tilde{\rho}^n \cdot \tilde{\mathbf{e}}_n. \end{aligned} \quad (1.3)$$

Then consider a point $X \in M$. The following formulas are obvious:

$$\overrightarrow{OX} = \overrightarrow{OO'} + \overrightarrow{O'X}, \quad \overrightarrow{O'X} = \overrightarrow{O'O} + \overrightarrow{OX},$$

By means of them we can find the relation of the coordinates of a point X in two different coordinate systems $O, \mathbf{e}_1, \dots, \mathbf{e}_n$ and $O', \tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$:

$$x^i = \rho^i + \sum_{j=1}^n S_j^i \tilde{x}^j, \quad \tilde{x}^i = \tilde{\rho}^i + \sum_{j=1}^n T_j^i x^j. \quad (1.4)$$

Though the vectors $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$ differ only in sign ($\tilde{\boldsymbol{\rho}} = -\boldsymbol{\rho}$), their coordinates in formulas (1.4) are much more different:

$$\rho^i = - \sum_{j=1}^n S_j^i \tilde{\rho}^j, \quad \tilde{\rho}^i = - \sum_{j=1}^n T_j^i \rho^j.$$

This happens because $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$ are expanded in two different bases (see the above expansions (1.3)).

The facts from the theory of affine spaces, which we stated above, show that considering affine spaces is a proper way for geometrization of the linear algebra.

A vector is an algebraic object: we can add vectors, we can multiply them by numbers, and we can form linear combinations of them. In affine space the concept of a point becomes paramount. Points form straight lines, planes, and their multidimensional generalizations — linear submanifolds. In affine spaces we have a quite natural concept of parallel translations and, hence, we can define the concept of parallelism for linear submanifolds. The geometry of two-dimensional affine spaces is called the *planimetry*, the geometry of three-dimensional affine spaces is called the *stereometry*. Affine spaces of higher dimensions are studied by a geometrical discipline which is called the *multidimensional geometry*.

§ 2. Euclidean point spaces. Quadrics in a Euclidean space.

DEFINITION 2.1. An affine space (M, V) over the field of real numbers $\mathbb{R}$ is called a *Euclidean point space* if the space V acting on M by parallel translations is equipped with a structure of a Euclidean vector space, i. e. if in V some positive quadratic form g is fixed.

In affine spaces, which we considered in previous section, a very important feature was lacking: there was no concept of a length and there was no concept of an angle. The structure of a Euclidean space given by a quadratic form g brings this lacking feature in. Let A and B be two points of a Euclidean point space M . They determine a vector $\mathbf{v} \in V$ specified by the condition $p_{\mathbf{v}}(A) = B$ (this vector is identified with the arrowhead segment $\overrightarrow{AB}$). The norm of the vector $\mathbf{v}$ determined by the quadratic form g is called the *length* of the segment AB or the *distance* between two points A and B : $|AB| = |\mathbf{v}| = \sqrt{g(\mathbf{v})}$. Due to the equality $|- \mathbf{v}| = |\mathbf{v}|$ we derive $|AB| = |BA|$.

Let $\overrightarrow{AB}$ and $\overrightarrow{CD}$ be two arrowhead segments in a Euclidean point space. They are geometric representations of two vectors $\mathbf{v}$ and $\mathbf{w}$ of V . The *angle* between $\overrightarrow{AB}$ and $\overrightarrow{CD}$ by definition is the angle between vectors $\mathbf{v}$ and $\mathbf{w}$ determined by the formula (1.6) of Chapter V.

DEFINITION 2.2. A coordinate system $O, \mathbf{e}_1, \dots, \mathbf{e}_n$ in a finite-dimensional Euclidean point space (M, V, g) is called a *rectangular Cartesian coordinate system* in M if $\mathbf{e}_1, \dots, \mathbf{e}_n$ is an orthonormal basis of the Euclidean vector space (V, g) .

DEFINITION 2.3. A *quadric* in a Euclidean point space M is a set of points in M whose coordinates $x^1, \dots, x^n$ in some rectangular Cartesian coordinate system $O, \mathbf{e}_1, \dots, \mathbf{e}_n$ satisfies some polynomial equation of degree two:

$$\sum_{i=1}^n \sum_{j=1}^n a_{ij} x^i x^j + 2 \sum_{i=1}^n b_i x^i + c = 0. \quad (2.1)$$

The definition of a quadric is not coordinate-free. It is formulated in terms of some rectangular Cartesian coordinate system $O, \mathbf{e}_1, \dots, \mathbf{e}_n$. However, passing to another Cartesian coordinate system is equivalent to a linear change of variables in the equation (2.1) (see formulas (1.4)). Such a change of variables changes the coefficients of the polynomial in (2.1), but it does not change the structure of this equation in whole. A quadric continues to be a quadric in any Cartesian coordinate system.

Let $O', \tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ be some other rectangular Cartesian coordinate system in M . Let's consider the passage from $O, \mathbf{e}_1, \dots, \mathbf{e}_n$ to $O', \tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$. In this case transition matrices S and T in (1.4) appear to be orthogonal matrices (see formulas (1.13) in Chapter V. We can calculate the coefficients of the equation of quadric in the new coordinate system. Substituting (1.4) into (2.1), we get

$$\tilde{a}_{qp} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} S_q^i S_p^j, \quad (2.2)$$

$$\tilde{b}_q = \sum_{i=1}^n b_i S_q^i + \sum_{i=1}^n \sum_{j=1}^n a_{ij} \rho^j S_q^i, \quad (2.3)$$

$$\tilde{c} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} \rho^i \rho^j + \sum_{i=1}^n b_i \rho^i + c. \quad (2.4)$$

Now the problem of bringing the equation of a quadric to a canonic form is formulated as the problem of finding a proper rectangular Cartesian coordinate system in which the equation (2.1) has the most simple canonic form.

The formula (2.2) coincides with the transformation formula for the components of a quadratic form under a change of basis (see (1.11) in Chapter IV). Hence, we conclude that each quadric in M is associated with some quadratic form in V . The form a determined by the matrix a_{ij} in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ is called the *primary quadratic form* of a quadric (2.1).

Let's consider the associated operator f_a determined by the primary quadratic form a (see formula (3.9) in Chapter V). The operator f_a is a selfadjoint operator in V ; it determines the expansion of the space V into the direct sum of two mutually orthogonal subspaces $\text{Ker } f_a$ and $\text{Im } f_a$:

$$V = \text{Ker } f_a \oplus \text{Im } f_a \quad (2.5)$$

(see (3.14) in Chapter V). The matrix of the operator f_a is given by the formula

$$F_j^i = \sum_{k=1}^n g^{ik} a_{kj}, \quad (2.6)$$

where g^{ik} is the matrix inverse to the Gram matrix of the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. Apart from f_a , we define a vector $\mathbf{b}$ through its coordinates given by formula

$$b^i = \sum_{k=1}^n g^{ik} b_k. \quad (2.7)$$

The definition of $\mathbf{b}$ through its coordinates (2.7) is essentially bound to the coordinate system $O, \mathbf{e}_1, \dots, \mathbf{e}_n$. This is because the formula (2.3) differs from the standard transformation formula for the coordinates of a covector under a change of basis (see (2.4) in Chapter III). Let's rewrite (2.3) in the following form:

$$\tilde{b}_q = \sum_{i=1}^n S_q^i \left(b_i + \sum_{j=1}^n a_{ij} \rho^j \right). \quad (2.8)$$

Then let's consider the expansion of the vector $\mathbf{b}$ into the sum of two vectors $\mathbf{b} = \mathbf{b}^{(1)} + \mathbf{b}^{(2)}$ according to the expansion (2.5) of the space V . This expansion induces the expansion $b_i = b_i^{(1)} + b_i^{(2)}$, where $b_i^{(1)}$ are transformed as follows:

$$\tilde{b}_q^{(1)} = \sum_{i=1}^n S_q^i b_i^{(1)}. \quad (2.9)$$

The vector $\mathbf{b}^{(2)}$ in the expansion $b = b^{(1)} + b^{(2)}$ can be annihilated at the expense of proper choice of the coordinate system. Let's determine the vector $\boldsymbol{\rho} = \overrightarrow{OO'}$ from the equality $\mathbf{b}^{(2)} = -f_a(\boldsymbol{\rho})$. Though it is not unique, the vector $\boldsymbol{\rho}$ satisfying this equality does exist since $\mathbf{b}^{(2)} \in \text{Im } f_a$. For its components we have

$$b_i^{(2)} + \sum_{j=1}^n a_{ij} \rho^j = 0, \quad (2.10)$$

this follows from $\mathbf{b}^{(2)} = -f_a(\boldsymbol{\rho})$ due to (2.6) and (2.7). Substituting (2.10) into (2.8), we get the following equalities in the new coordinate system:

$$\tilde{\mathbf{b}}^{(2)} = \mathbf{0}, \quad \tilde{\mathbf{b}} = \tilde{\mathbf{b}}^{(1)}.$$

The relationships (2.9) show that the numbers $b_i^{(1)}$ cannot be annihilated (unless they are equal to zero from the very beginning). These numbers determine the vector $\mathbf{b}^{(1)} \in \text{Ker } f_a$ which does not depend on the choice of a coordinate system. As a result we have proved the following theorem.

THEOREM 2.1. *Any quadric in a Euclidean point space (M, V, g) is associated with some selfadjoint operator f and some vector $\mathbf{b} \in \text{Ker } f$ such that in some rectangular Cartesian coordinate system the radius vector $\mathbf{r}$ of an arbitrary point of this quadric satisfies the following equation:*

$$(f(\mathbf{r}) | \mathbf{r}) + 2(\mathbf{b} | \mathbf{r}) + c = 0. \quad (2.11)$$

The operator f determines the leading part of the equation (2.11). By means of this operator we subdivide all quadrics into two basic types:

- (1) *non-degenerate quadrics*, when $\text{Ker } f = \{\mathbf{0}\}$;
- (2) *degenerate quadrics*, when $\text{Ker } f \neq \{\mathbf{0}\}$.

For non-degenerate quadrics the vector $\mathbf{b}$ in (2.11) is equal to zero. Therefore, non-degenerate quadrics are subdivided into three types:

- (1) *elliptic type*, when $c \neq 0$ and the quadratic form $a(\mathbf{x}) = (f(\mathbf{x}) | \mathbf{x})$ is positive or negative, i. e. can be made positive by changing the sign of f ;
- (2) *hyperbolic type*, when $c \neq 0$ and the quadratic form $a(\mathbf{x}) = (f(\mathbf{x}) | \mathbf{x})$ is not sign-definite, i. e. its signature has both pluses and minuses;
- (3) *conic type*, when $c = 0$.

Degenerate quadrics are subdivided into two types:

- (1) *parabolic type*, when $\dim \text{Ker } f = 1$ and $b \neq 0$;
- (2) *cylindric type*, when $\dim \text{Ker } f > 1$ or $b = 0$.

The equation (2.1) in the case of a non-degenerate quadric of **elliptic** type can be brought to the following canonic form:

$$\frac{(x^1)^2}{(a_1)^2} + \dots + \frac{(x^n)^2}{(a_n)^2} = \pm 1.$$

This is the canonic equation of a non-degenerate quadric of **hyperbolic** type:

$$\frac{(x^1)^2}{(a_1)^2} \pm \dots \pm \frac{(x^n)^2}{(a_n)^2} = \pm 1.$$

The canonic equation of a non-degenerate quadric of **conic** type is homogeneous:

$$\frac{(x^1)^2}{(a_1)^2} \pm \dots \pm \frac{(x^n)^2}{(a_n)^2} = 0.$$

The equation (2.1) in the case of a degenerate quadric of **parabolic** type can be brought to the following canonic form:

$$\frac{(x^1)^2}{(a_1)^2} \pm \dots \pm \frac{(x^{n-1})^2}{(a_{n-1})^2} = 2x^n.$$

If $n = \dim M > 1$, then in a canonic equation of a quadric of **cylindric** type there is no explicit entry of at least one variable. Therefore, we can reduce the dimension of the space M . The reduced quadric can belong to any one of the above four types. If it is again of cylindric type, then we can repeat the reduction procedure. This process can terminate in some intermediate dimension yielding the reduced quadric of some non-cylindric type. Otherwise we shall reach the dimension $\dim M = 1$. In one-dimensional Euclidean point space there is no quadrics of cylindric type. Therefore, the quadrics of **cylindric** type are those which belong to one of the non-cylindric types in the reduced dimension.

REFERENCES.

1. Kurosh A. G. *Course of general algebra*, «Nauka» publishers, Moscow.
2. Sharipov R. A. *Course of differential geometry*¹, Bashkir State University, Ufa, 1996; see online [math/0412421](http://arXiv.org/math/0412421) in Electronic Archive <http://arXiv.org> and r-sharipov/r4-b3.htm in GeoCities.
3. Sharipov R. A. *Classical electrodynamics and the theory of relativity*, Bashkir State University, Ufa, 1997; see online [physics/0311011](http://arXiv.org/physics/0311011) in Electronic Archive <http://arXiv.org> and r-sharipov/r4-b5.htm in GeoCities.
4. Kostrikin A. I. *Introduction to algebra*, «Nauka» publishers, Moscow, 1977.
5. Beklemishev D. V. *Course of analytical geometry and linear algebra*, «Nauka» publishers, Moscow, 1985.
6. Kudryavtsev L. D. *Course of mathematical analysis, Vol. I and II*, «Visshaya Shkola» publishers, Moscow, 1985.
7. Sharipov R. A. *Quick introduction to tensor analysis*, free online publication [math.HO/0403252](http://arXiv.org/math.HO/0403252) in Electronic Archive <http://arXiv.org>, 2004; see also r-sharipov/r4-b6.htm in GeoCities.

¹ The references [2] and [3] are added in 1998, the reference [7] is added in 2004.